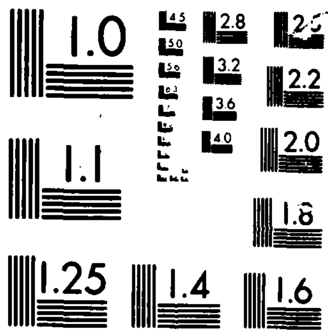
AD-A169 014 LEARNING RESEARCH LABORATORY: PROPOSED RESEARCH ISSUES 1/2
(U) UNIVERSAL ENERGY SYSTEMS INC DAYTON OH
I L GOLDSTEIN ET AL. JUN 86 AFHRL-TP-85-54

UNCLASSIFIED F33615-82-C-0006 F/G 5/10 NL


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.8
3.2
3.6
4.0
4.5
5.0

AFHRL-TP-85-54

AIR FORCE

HUMAN RESOURCES LABORATORY

AD-A169 014

DTIC FILE COPY

LEARNING RESEARCH LABORATORY: PROPOSED RESEARCH ISSUES

Irwin L. Goldstein
University of Maryland
College Park, Maryland 20742

Robert M. Gagné
Florida State University
Tallahassee, Florida 32306

Robert Glaser
University of Pittsburgh
Pittsburgh, Pennsylvania 15260

James M. Royer
University of Massachusetts
Amherst, Massachusetts 01003

Thomas J. Shuell
State University of New York at Buffalo
Amherst, New York 14260

David L. Payne, Lt Col, USAF

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

June 1986
Final Paper for Period March 1983 - March 1984

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

86 7 7 021

NOTICE

WILLIAM E. ALLEY, Scientific Advisor
Manpower and Personnel Division

RONALD L. KERCHNER, Colonel, USAF
Chief, Manpower and Personnel Division

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

AD-A169014

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFHRL-TP-85-54
6a. NAME OF PERFORMING ORGANIZATION Universal Energy Systems Inc. 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Manpower and Personnel Division
6c. ADDRESS (City, State, and ZIP Code) 4401 Dayton-Zenia Road Dayton, Ohio 45432	7b. ADDRESS (City, State, and ZIP Code) Air Force Human Resources Laboratory Brooks Air Force Base, Texas 78235-5601
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Air Force Human Resources Laboratory 8b. OFFICE SYMBOL (If applicable) HQ AFHRL	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F33615-82-C-0006
8c. ADDRESS (City, State, and ZIP Code) Brooks Air Force Base, Texas 78235-5601	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
62205F	7710	00	08
61102F	2313	T1	33

11. TITLE (Include Security Classification)

Learning Research Laboratory: Proposed Research Issues

12. PERSONAL AUTHOR(S)

Goldstein, I. L.; Gagne, R. M.; Glaser, R.; Royer, J. M.; Shuell, T. J.; Payne, D. L.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 14Mar83 TO 30Mar84	June 1986	98

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP
05	09	
05	10	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

cognitive theory
comprehension processes
information processing
instructional design
learning theory
memory processes (Continued)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

In the interest of planning a laboratory to conduct basic research in learning and training pertinent to the training functions of the Air Force, the Air Force Human Resources Laboratory assembled a panel of nationally recognized experts in instructional psychology. Each expert presented his view of contemporary research and theory in instructional psychology and identified a set of research issues felt to be critical for the refinement of training procedures in the Air Force.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Nancy A. Perrigo, Chief, STINFO Office	22b. TELEPHONE (Include Area Code) (512) 536-3877 22c. OFFICE SYMBOL AFHRL/TSR

Item 18 (Concluded):

skill acquisition
skill retention
training

AFHRL Technical Paper 85-54 June 1986

LEARNING RESEARCH LABORATORY: PROPOSED RESEARCH ISSUES

Irwin L. Goldstein
University of Maryland
College Park, Maryland 20742

Robert M. Gagné
Florida State University
Tallahassee, Florida 32306

Robert Glaser
University of Pittsburgh
Pittsburgh, Pennsylvania 15260

James M. Royer
University of Massachusetts
Amherst, Massachusetts 01003

Thomas J. Shuell
State University of New York at Buffalo
Amherst, New York 14260

David L. Payne, Lt Col, USAF

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

Reviewed by

William E. Alley, Scientific Advisor
Manpower and Personnel Division

Submitted for publication by

Ronald L. Kerchner, Colonel, USAF
Chief, Manpower and Personnel Division

This publication is primarily a working paper. It is published solely to document work performed.

SUMMARY

This technical paper documents the proceedings of a workshop that was convened to formulate the technical program for the establishment of a learning research facility at the Air Force Human Resources Laboratory. Attendees at the workshop included nationally recognized experts in instructional psychology from academia and from laboratories of the other Military Services. These participants were asked to review the needs of the Air Force, compare those needs with relevant technology, and propose approaches for a productive research program. Each paper presents the author's view of contemporary research and theory in instructional psychology and identifies a set of research issues believed to be critical for the refinement of training procedures in the Air Force. ~~The research issues~~ identified provide an excellent foundation for the development of new research initiatives in the learning and training areas.

PREFACE

This technical paper documents an initiative by the Air Force Human Resources Laboratory to conduct basic research on learning that could be used in the training of Air Force personnel. The technical direction for the initiative was provided by Lt Col David L. Payne, who inspired the efforts that produced this document. Dr. Robert M. Gagné was also a very significant contributor. During most of this effort, he was serving at the Air Force Human Resources Laboratory (AFHRL) as an expert consultant while on leave from Florida State University. He graciously continued to contribute to the effort after returning to the University.

This effort represents a portion of the technical program of the Manpower and Personnel Division of the Air Force Human Resources Laboratory (AFHRL/MO). This effort also was pertinent to the technical program of the AFHRL Logistics and Human Factors Division (AFHRL/LR). This effort was performed as Task 09, Workshop on Instructional Technology, under Contract F33615-82-C-0006 (SB5448-82-C-0099) with Universal Energy Systems, Inc. This contract is documented under Work Unit 1710-00-08, Contributive Research in Logistics Factors and Training Techniques, with Mr. Craig Cassino serving as Work Unit Monitor. Mr. Edward Boyle was the technical representative of AFHRL/LR on this effort. For this effort, Drs. Goldstein, Glaser, Royer, and Shuell served as employees of Universal Energy Systems, Inc., under the general supervision of Dr. Ross L. Morgan. Dr. Goldstein served as Chairman of the group. In addition to authoring two chapters of the present report, Dr. Goldstein coordinated the efforts of all of the authors.

Significant technical contributions to this effort were made by individuals who met with the authors and briefed them on relevant research and issues. These individuals include the following: Dr. Raymond E. Christal and Dr. Hendrick W. Ruck of AFHRL/MO; Dr. Joseph Y. Yasutake of the AFHRL Training Systems Division; Dr. William E. Montague of the Navy Personnel Research and Development Center; and Dr. Joseph S. Ward of the US Army Research Institute for the Behavioral and Social Sciences. These individuals also reviewed drafts of the present document, as did Dr. William C. Tirre, AFHRL/MO, who provided substantive review of each paper. In addition to those whose contributions have been acknowledged, many unnamed individuals served commendably on the team that produced this technical paper. It is hoped that appreciation for their services has been acknowledged appropriately at other times and places.

TABLE OF CONTENTS

Table of Contents (Continued)

Table of Contents (Continued)

Table of Contents (Concluded)

LIST OF FIGURES

LEARNING RESEARCH LABORATORY: PROPOSED RESEARCH ISSUES

CHAPTER 1

INTRODUCTION

The Air Force Human Resources Laboratory (AFHRL), is preparing plans for a learning research laboratory that can accomplish generalizable research on problems of training common to many Air Force jobs. The purpose of such a laboratory would be to conduct basic research, in learning and training, that would have implications for the training functions of the Air Force.

To help formulate the technical program of a learning research laboratory, several noted researchers were asked to review the pertinent needs of the Air Force, compare those needs with the relevant technology, and propose approaches to a productive research program. As a first step toward the basic objective, a workshop was conducted. This workshop was attended by the authors of the present document and representatives of AFHRL, the Navy Personnel Research and Development Center (NPRDC), and the Army Research Institute for the Behavioral and Social Sciences (ARI). During the workshop, representatives of AFHRL presented information about the Air Force training process and related research programs in AFHRL. The representatives of NPRDC and ARI provided summaries of their research programs and offered useful advice on further research.

Following these presentations, the authors presented preliminary versions of their papers. Their presentations had the following objectives:

1. State the author's view of contemporary research and theory on cognitive learning and individual differences in cognitive processes that are relevant to the planning of the laboratory and its program.

2. Identify the research problems in the area of basic learning and training research that are considered to be promising in terms of valuable outcomes.

3. Indicate the applicability of the research areas to problems of Air Force training.

Each of the papers was followed by discussions among all of the participants. Following the workshop, the authors developed formal papers. Those papers were circulated for comments to all individuals who attended the workshop. The comments also were circulated to all participants. After this process, revised papers were developed and are presented here. The final chapter provides a summary of the research issues that are discussed in the various chapters.

The participants in this effort hope that the research ideas contained in these papers will contribute to a productive research agenda for an Air Force learning research laboratory.

CHAPTER 2

GENERAL ORIENTATION TO ESTABLISHMENT OF A LEARNING/TRAINING RESEARCH LABORATORY

David L. Payne
Air Force Human Resources Laboratory

Robert M. Gagné
Florida State University

The Air Force has 223 enlisted specialties in 48 career fields and has 218 officer specialties. Training is conducted by the Air Training Command (ATC). For enlisted airmen, a period of 6 weeks of basic training is usually followed by assignment to a school where they are taught the knowledge and skills relevant to a particular type of job assignment. Schools teach courses in each of the broad career fields, such as aircraft mechanic, communications systems maintenance, personnel specialist, and accounting and finance. In addition to resident training, operational field training and on-the-job training (OJT) may be given in order to refresh skills infrequently used, or to meet new requirements because of the introduction of system changes.

The Need for a Learning Research Laboratory

The need for a laboratory to investigate problems of learning arises due to problems associated with conducting certain types of research in operational training environments. In the Air Force, resident training courses are highly structured to present information required to bring the student's knowledge up to predetermined levels of acceptability. The courses are designed to do this in the most efficient manner possible, with production quotas dictated by the needs of the Air Force. These operational requirements often make it extremely difficult to conduct research which requires manipulations in methods of instruction, course length, and related factors of interest. In this setting, we often must forsake certain requirements of appropriate research design (for example, use of control groups) to avoid disruptions in the ongoing training system. Also, changes in training materials, curricula, media and methodologies are introduced into the operational setting on a periodic basis. These changes are often not properly evaluated in an experimental setting prior to implementation, due to constraints in the operational environment.

This is not to say that the Air Force is unwilling to disrupt their day-to-day training regimen if the research holds great promise of operational benefits; indeed, this has happened in previous studies in pilot selection, computer-assisted instruction, and other areas. However, demonstrating "great promise" on an a priori basis for training research is often not possible for basic research efforts. As part of the AFHRL learning abilities research to be described by Dr. Christal, we will identify individual differences in abilities that will generate hypotheses to be tested by research in technical training. However, it may well be difficult to convince the operational trainers that the test of a theory is worth the disruption of ongoing training activities, the cost of developing two courses, and so on. A research laboratory, separate from normal training operations, would provide the opportunity to conduct limited tests in a far less disruptive and less costly manner.

Areas of Research

What kinds of research on learning and instruction need to be included in the program we are planning?

I'm sure you don't expect me to provide any specific answers to that question, because that's what we're asking you to give us in the presentations you make and the discussions that grow out of them. Here, though, are some very general areas that appear to contain the kinds of problems faced by Air Force training:

1. Studies need to be conducted pertaining to the acquisition, retention, and reacquisition of Air Force job-relevant skills and knowledges. Without highly controlled experimental conditions, acquisition rates are extremely difficult to measure. The Air Force now trains to certain percentage standards of performance, with no examination of the effects of individual differences in achieving these levels of performance. One of the problems in forecasting knowledge and skill acquisition rates is that there may be interactions between aptitude and training methods. For example, some individuals may learn faster from auditory presentations than from visual presentation of training materials. Other questions of interest in this domain include the following: Can long-term retention be predicted from indicators of short-term learning or of immediate recall? If transfer depends on similarity or on identity of elements, what are the elements? Is the schema a useful way to conceptualize the elements that need to be similar? What are the effects of continued practice, and what is accomplished when a skill is automatized?

2. There have been various studies indicating a relationship between aptitude and training time. However, at the present time, no one can express aptitude-training time tradeoffs with any degree of precision. If one were to raise the aptitude level of individual groups assigned to a training course by X points, what savings in training time could be realized in bringing personnel to the same level of competence? If the quality of recruits decreases, as indicated by lower aptitude scores, how much additional training time will be required to achieve the same level of competence? Are there some courses in which an acceptable level of competence cannot be attained with lower-aptitude accessions? Questions of this sort suggest a need to undertake some fundamental investigations of "time to learn" and how it is affected by the nature of the skill being trained and by the aptitude of the learner.

3. A critical need exists to provide an appropriate research setting in which job performance criteria can be used to validate current and proposed ability tests. Although most managers accept that our current selection tests are reasonably valid for predicting school success, they often ask the question, "How well do the tests predict performance on the job?" This is a fair question which the research community has not adequately answered, primarily due to technical problems associated with using field job performance as a validation criterion. These problems have proven so severe that many of us doubt they will ever be overcome without major changes to our current validation strategies. I suggest that a primary objective of the proposed research laboratory should be to collect reliable performance data for test validation purposes, thus providing reliable "intermediate criteria."

Representative Tasks

In order to have a good basis for the research in such a laboratory, we intend to select and use some reference tasks representative of what people must learn to do in Air Force jobs. Dr. Gagné has spent some time searching for components of some Air Force enlisted specialties that could serve as reference tasks, to be learned over a time span of several days or less. He is trying to find the kinds of job components that would require dealing with the following kinds of things: (a) maintaining electrical systems; (b) maintaining mechnical systems; (c) computing with numbers; (d) composing written messages; and (e) sorting, classifying, and coding information on forms.

The idea would be to select some convenient job components from these specialties or from others like them. These components (not necessarily single tasks) would be taught to airman recruits over a reasonably short period of time and under laboratory control of conditions. These would then become the reference tasks to serve as criteria for performance effects of instructional variables in experimental studies.

It is often a difficult part of learning research to select learning tasks that represent "the real world." Investigators often have to think hard to devise tasks that will simulate problems or other learning situations that are realistic. Our approach is to try to select and present pieces of actual jobs. This seems to us a part of the answer of how to begin. Other suggestions, we hope, will be coming from this workshop.

CHAPTER 3

RESEARCH ON LEARNING AND RETAINING SKILLS

Robert M. Gagné
Florida State University

Having in mind the need for research that will shed light on the conditions of skill learning and skill maintenance, as it might be conducted in an Air Force laboratory, I have prepared the following account of what seems to me to be promising lines and directions for a research program. This begins with an account of a set of "benchmark" learning tasks, derived directly from Air Force jobs, and intended to serve as representatives of those jobs. Following this are descriptions of lines of research on skill learning, mental effort as an indicator of task difficulty, skill retention, and relearning. Description of each of these major areas is accompanied by a list of references, collected by area, at the end of this paper.

Benchmark Tasks for Learning Research

Frequently, tasks chosen for studies of learning in military settings cover a range from fairly simple, real tasks to more complex tasks that are designed to simulate actual military tasks, and include learning tasks designed to incorporate the cognitive processes of actual tasks. An example of a comprehensive study of this sort is Fox, Taylor, and Caylor (1969). Tasks used by these investigators included the actual task of rifle assembly and disassembly and a simulated task, called "missile preparation," that required the learning of a 34-step procedure in responding to controls on a panel. In more recent Army work, learning and retention of a number of actual job-tasks have been studied in field-type research. For example, Goldberg, Drillings, and Dressel (1981) used the actual task of boresighting and zeroing the M60-A1 main gun. Hagman (1980) investigated several parts of the job of charging system output using the 500A Sun Test Stand generator; and Holmgren, Hilligoss, Swezey, and Eakins (1979) looked at a variety of tasks oriented around the machine gun, the squad radio, tube artillery, and radar equipment.

Selection of Job Components

There are, then, different approaches to the selection of representative job-tasks, which are therefore valid in the sense of possessing verisimilitude as judged by observers who are informed and unbiased. Obviously the most direct method is to select a group of tasks that are integrally related to some coherent job operation. This organized set of tasks may be called a job component. Sometimes a job component of this sort may be revealed by the organized set of tasks called a "duty." But this is not always the case, and direct observation of job incumbents may be necessary to identify appropriate job components.

Once selected, the representativeness of a job component may be assessed by means of various measures of its tasks available through the Comprehensive Occupational Data Analysis Programs (CODAP) system. These include the percentage of personnel performing the tasks and the rated training emphasis. The job component selected should contain, as a minimum, tasks which have a high percentage performance by incumbents of the Air Force Specialty (AFS). Otherwise, it seems appropriate that the job components represent a range of different AFSs, in the following manner:

1. A procedure of 10 or more steps outlined in a technical manual pertaining to checking the operation of complex equipment (possibility: AFSC 423X5, Aerospace Ground Equipment (AGE) Mechanic).

2. A procedure of six or seven steps to be recalled without the aid of printed text, involving removing or installing mechanical components (possibility: AFSC 423X5, AGE Mechanic).

3. The inspection and calibration of a test set, requiring responses to a number of different dials and indicators (10 to 16) on a panel (possibility: 423X0, Aircraft Electrical Systems Specialist).

4. The composition of relatively brief, partially structured messages in memoranda and correspondence (possibility: 732X0, Personnel Specialist).

5. The carrying out of "paper" procedures, such as disbursement and accounting (possibility: 672X2, Financial Service).

6. The computation of trends and other numerical measures (possibility: 691X0, Cost and Management Analysis Specialist).

Validating Aptitude Tests and Batteries

The representative nature of job components means that they can play the role of intermediate criteria against which aptitude tests can be validated. This potential use serves to confirm the desirability of a suitable range of job components in different aptitude areas. Referring to the previously described possibilities, the Electronics area is represented by (1) and (3), Mechanical is represented by (2), General is represented by (4) and (6), and Administrative is represented by (5). In view of the current difficulties and suspected inappropriateness of school grades as criteria for Armed Services Vocational Aptitude Battery (ASVAB) and other tests, this particular use of job components in a laboratory setting may be seen as a requirement with high priority. Correlations would be obtained between job component performance scores and the various sub-scores of ASVAB, as well as with individual tests of aptitude.

Validating Job Difficulty Measures

CODAP data include ratings of task difficulty that are obtained by asking supervisors "How long does it take to learn the task?" This measure needs to be validated by comparing it with a direct measure of "time to learn" on the same task. Obtaining the latter measure on a job component will require some methodological development. Time to learn has not been widely employed as a measure, and some arbitrary criteria of learning and performance may have to be set before meaningful research can be done. A direct measure of time to learn for job components would make possible interesting and valuable comparisons with task difficulty as currently assessed.

Validating Learning Measures

In a learning research program, it may often be convenient to present the learning task to be investigated by means of a computer with TV monitor, TV disc, or similar device. This means that the task will likely be an abstraction from the "real" task. Yet, such abstractly derived tasks may be necessary to employ for the study of such variables as mental effort, varieties of practice, overlearning, and automatization. Sometimes, research findings may make possible more or less direct applicability to the planning of training. In many instances, however, the findings of laboratory-type studies on synthetic tasks may be checked against comparable findings with real job-tasks, that is, with job components.

Validating Cognitive-Skill Analysis

With the purpose of defining training requirements, a number of current projects of the field-research type propose to analyze tasks of Air Force jobs. There are two outcomes of this kind of analysis. First, the analysis reveals the supporting skills and knowledges of a task. According to Instructional Systems Development (ISD) procedures, these specific skills and knowledges are incorporated into plans for training and into course design. Examination of the task, together with its supporting skills and knowledges, may also yield a characterization of the task as partaking of one or more critical cognitive processes (such as high mental effort). This second kind of outcome of analysis is likely to reveal some requirements for classification testing.

Supporting skills and knowledges of a task may be validated by studies having the framework of transfer of training. Students who learn the identified skills and knowledges should learn to perform the total task more readily (faster or with fewer errors) than students who have not learned the same skills and knowledges. Since it is extremely difficult to perform such studies in the field, basic research investigations that validate this type of "component analysis" of job-tasks would appear to be of substantial value. Of course, the most important validation is of the method and not of the specific skills and knowledges identified.

The relation of particular cognitive processes (or skills) to task performance can also be validated by correlations between basic cognitive skills revealed by analysis and performance on job components. This is, of course, simply a particular instance of aptitude test validation as previously described.

Establishing Reliable Measures of Performance

Since job components are considered to play such a crucial role in validation of a variety of testing and learning operations, it will be a necessary goal to select and devise measures of performance that are as reliable as possible. Low reliability has often been an obstacle to the use of job performance measures. Can reasonably high reliability be obtained with measures of performance on job components?

The nature of measures applicable to job components cannot be estimated ahead of time with any assurance. Probably, the following characteristics ought to be aimed for:

1. Criterion-referenced, performance-based.

2. Yielding a range of scores reflecting performance accuracy.

3. Having reliability that is conceived as coefficient of equivalence (Siegel et al., 1979); i.e., correlation between measures obtained from equivalent instruments.

Validating Methodology for Performance Assessment

The requirement for objective measures of job performance is widespread among projects dealing with selection and classification testing, training development and evaluation, and the structuring and revision of AFS categories. Many difficulties attend the attempts to collect data on the job performance of airmen assigned to various occupational specialties. For one thing, following initial school training, personnel in any single career field become widely scattered at operational bases in the United States and overseas; the result is a thinning of the ranks of airmen who could be considered to constitute groups of comparable individuals for study purposes. Although they may be roughly comparable at the end of first-school training, airmen in a single AFS encounter various job assignments and job conditions in different commands and different locations.

A second major consideration regarding the assessment of performance of airmen in such specialties as equipment maintenance is that it becomes an enormously complex task involving real equipment, parts of real equipment, simulation equipment, training devices, and other materials. Because of the necessity of employing these varieties of equipment and establishing standard conditions of observation, the whole effort becomes quite expensive. Examples of what may be required are contained in reports by Foley (1974); Shriver and Foley (1974); and Shriver, Hayes, and Hufhand (1975).

It would appear that widespread evaluation of job performance at a selection of Air Force bases is prohibitively expensive, as well as being disruptive of ongoing operations. Accordingly, it would seem desirable to validate methods of job performance assessment in the restricted, controlled environment of a laboratory. Thoroughgoing procedures such as those described by Shriver, Hayes, and Hufhand (1975) could be established in connection with a relatively minor system or equipment. Following this, methods of simplifying the procedures of performance assessment could be explored and tested. The aim would be a set of procedures that are practical for on-the-job performance assessment and that would yield valid and reliable measures.

Skill Learning Beyond Initial Acquisition

Observations of human performance that demonstrates intellectual skills often raise questions about the adequacy of the learning experience that has produced those skills. Skills that are not constantly used are often forgotten. Skills that are employed with moderate frequency may have to be subjected to a self-review process before being used, rather than being readily accessible for immediate application to each new specific situation. These circumstances suggest questions about the completeness or adequacy of learning and their effects on skill maintenance and generalizability.

Cognitive psychologists have suggested that skills may progress through a number of stages of adequacy beyond the point of initial acquisition. Neves and Anderson (1981) proposed that proceduralization of a skill begins with a transition from knowledge that is declarative information (evidenced by the learner's verbal statement) to knowledge that is procedural in the sense of applicability to one or more specific instances. Changes that go beyond initial acquisition of the procedure and that may take place either simultaneously or successively include the following:

1. Composition, in which certain parts of the total procedure are consolidated in such a way that "short-circuiting" of the procedure takes place.

2. Speed-up, which occurs as a result of composition and when certain unessential part-activities drop out with continued practice.

3. Automaticity, which makes possible parallel rather than successive processing of certain activities, such as the identification of appropriate conditions for employing a skill. Automization results in a performance that does not interfere with another task concurrently carried on (Shiffrin & Dumais, 1981).

Changes in learned capabilities which go beyond initial acquisition are described by Rumelhart and Norman (1978) as "tuning." As applied to schemata, the idea of tuning may include changes that accomplish any of the following: (a) improve accuracy, (b) generalize applicability, (c) specialize applicability, and (d) determine the "default values," or the characteristics of the typical case. Generally speaking, the changes in capability described by Neves and Anderson (1981) are expected to result from additional practice. The tuning changes may also be brought about with practice, but there may be other ways of effecting these changes following the initial attainment of the skill. The effects of practice may serve to amplify "resistance to forgetting" as Wickelgren (1972) suggests.

Research on Learning Beyond Initial Acquisition

Some areas of research that appear to be worthy of investigation include the following:

1. Can phases of learning beyond initial acquisition be identified, and can the operations for detecting them be specified? In a procedural task of 10 to 12 steps, learning could first be carried to the criterion of one perfect performance, and the number of trials recorded. An average number of trials would be obtained for groups in aptitude categories 1, 2, and 3. In the same categories, trials would be given to represent 150% and 200% "overlearning." Observations would be made to detect "composition" and "speed-up" of parts of the procedure. For example, steps that followed each other by good cueing might be expected to exhibit composition and speed-up, whereas steps that did not follow inherent cues might not behave in the same way. Composition and speed-up would be detected by the timing of part-activities and by other methods. Automatization would be detected by the degree of degradation in a secondary task (monitoring lights, or the like).

2. Can the phases of advanced skill compilation, as mentioned in 1, be attained by methods other than continued practice? As contrasted with "straight practice," alternative procedures to achieve composition, speed-up, and automatization might be (a) specific training on identification of the cues leading from one step to the next, (b) setting speed goals for the learner, (c) practice of the procedure in the presence of a secondary task (cf. Naylor, Briggs, & Reed, 1968), or (d) various combinations of these approaches. In general, the purpose of such studies would be to compare the effects of alternative methods with those of straight practice. It is possible that such methods might save time or otherwise accomplish the goal of skill consolidation more efficiently than would continued practice.

3. Proceeding to investigate further these advanced phases of learning, the questions include the following:

a. What are the effects of the composition phase on retention of a skill, relearning of the skill, and transfer of the skill to related tasks?

b. What are the effects of the composition phase plus the speed-up phase of advanced learning on retention of the skill, relearning of the skill, and transfer to related tasks?

c. What are the effects of automatization on retention, relearning, and transfer?

4. Automatization is a particularly interesting and probably important phase of advanced learning. What evidence can be obtained from parallel processing with automatized skills? This question can perhaps be answered by determining whether differences exist in automatization for different types of tasks (e.g., procedure vs. organized motor) and possibly for different numbers of tasks conducted simultaneously.

Measuring Task Difficulty as Mental Effort

The question of difficulty of the various tasks making up Air Force jobs and occupations is one which affects the planning of training and personnel management. Tasks of higher difficulty are assigned to advanced levels of AFSs and are usually assumed to require different patterns of aptitude predictors from those associated with lower difficulty. Differential task difficulty also influences the planning of training, in terms of durations of instructional modules and courses. In these various uses, "task difficulty" is a variable derived from estimations by on-the-job supervisors. Added precision may be expected from the development of direct measures of task difficulty, derived from observations of performance. Such a method would make possible validation of subjective estimates of task difficulty.

A number of investigators of human learning and performance accept the idea that human tasks vary in the proportion of limited-capacity central processing they demand (Johnston & Uhl, 1976; Kahneman, 1973; Moray, 1967; Tyler, Hertel, McCallum, & Ellis, 1979). The central processing is often referred to as "attention," and the allocation of the resource as "division of attention" (Shulman, Greenberg, & Martin, 1971). An individual may need to divide attention when required to perform a primary and a secondary task at one time. The degree to which the secondary task performance is diminished is often taken as a measure of the attentional demands of the primary task (Kerr, 1973).

As an example, Tyler et al. (1979) studied attentional demands, which they called "cognitive effort," in two primary tasks. One task was anagrams, in which, for example, the word "doctor" was transformed into "dortoc" (low effort) and "croodt" (high effort). A second primary task used sentences to be completed: "The girl was awakened by the frightening __________ (low effort)"; and "The man was alarmed by the frightening ________ (high effort)." The secondary task used in this study required the individual to respond as rapidly as possible by turning off a tone delivered through headphones (simple reaction time). The tone came on randomly at four different intervals following the presentation of each item of the primary task. Findings of the study were that reaction time was greater for high effort than for low effort tasks.

Measures of Task Difficulty

The study by Tyler et al. (1979), as well as other studies (Kerr, 1973; Posner & Boies, 1971), implies that the secondary task of reaction time to tone can be employed as a measure of "mental effort," and therefore, by implication, as a measure of task difficulty. Presumably, such a measure answers the question, "How much (what proportion of) attentional resource must be devoted to the performance of this task?"

It may be noted that the use of a reaction-time measure of mental effort is most appropriate for tasks that can be presented item by item (as in word recognition, numerical calculation, concept identification). Such a measure would not be appropriate for a task having a serial nature in which attention had to be maintained over a period of time during which several different kinds of processing, or several successive steps of processing, were being carried out. Johnston, Greenhag, Fisher, and Martin (1970) employed continuous compensatory tracking of an oscilloscope signal as a secondary task; their primary task was a list of words to be categorized. In this case, tracking error was used as the measure of the attentional "difficulty" of the primary task.

In the case of a procedural task, such as "adjusting a crystal oscillator," there are several steps which, although performed one by one, constitute a coherent sequential performance. It is conceivable that each step might be accompanied by a secondary reaction-time task, yielding individual subtask measures that might be summed to provide a total task measure of mental effort. Alternatively, if a simulated version of the task was being presented on a display screen, a tracking task could be employed in a secondary role to provide a tracking error score as a measure of mental effort. Independent of the particular measure employed, it is surely appropriate to note the factor of number of steps (cf. Shields, Goldberg, & Dressel, 1979) as a contributor to task difficulty.

Research Questions

1. For a procedural task of 4 to 7 steps to be carried out without an external memory aid (e.g., putting an electronic device into operation), what is the mental effort at three different stages of task practice? It may be supposed that when the task is first done, mental effort may be high. After a few trials of practice, the effort may be reduced and after still more practice, reduced further. Using the reaction-time technique of secondary task measurement, a tone could be introduced after the beginning of each step of the task, perhaps using random intervals in the manner of Tyler et al. (1979). Such a method would relate degree of training with mental effort. In addition, this method would yield a measure of the difficulty of each step, data which may contribute to an understanding of "difficulty" as applied to the task as a whole.

2. For the same type of task, is difficulty assessed by mental effort included in, or independent of, a measure of task difficulty which is "time to learn"? Time to learn could, of course, be measured by recording either (a) number of trials or (b) elapsed time required to reach some specified criterion of performance.

3. For the same type of task, what relations exist among the measures of difficulty called (a) mental effort, (b) time or trials to learn, and (c) rated difficulty as given by the CODAP system?

4. For a type of task requiring the learning and differentiation of a number of concepts, the questions posed in 1, 2, and 3 may be asked. Examples of this kind of task are the military map symbols employed by Fox, Taylor, and Caylor (1969) and the Power Station Operator Task used in a current AFHRL effort. Alternatively, it should be possible to identify a component of an Air Force Specialty Code (AFSC) which could be represented in a laboratory setting, either directly or by simulation. (See the section on "Benchmark Tasks for Learning Research.")

5. Another question of importance is the relation between one or more measures of mental effort and amount of retention over the interval of a week or more. A summary of recent research on this problem is given by Rose, McLaughlin, and Felker (1981) who point out the limitations of these studies.

6. Discovering how these different measures of mental effort relate to each other in learning and in retention should make possible a choice of a measure to represent "task

difficulty." The application of the measures to at least two different kinds of tasks (1 and 4) will make a beginning in answering the question of generalizability. If a "best choice" of task difficulty measures cannot be made, an understanding of how these measures are related will be of interest.

Skill Maintenance and Skill Forgetting

Maintaining the quality of learned skills in military jobs is an objective that forms the basis for a number of research efforts over the past 10 years or so (Arima & Neil, 1978; Goldberg, Drillings, & Dressel, 1981; Prophet, 1976; Schendel & Hagman, 1980). Concern with the forgetting of skills arises from a number of circumstances, including the following:

1. Some of the tasks on which training is given may not be used on the job immediately and tend to suffer forgetting.

2. Skills learned initially may not be required until 5-level (journeyman-level) job proficiency must be demonstrated and may be forgotten during that interval.

3. When training courses are of lengthy duration, skills learned in early weeks may be forgotten by the end of course.

Forgetting of skills must be deemed an unfortunate occurrence in any of these situations and a factor which is frequently detrimental to job performance. The avoidance of skill forgetting and the adoption of procedures to maintain skills would appear to be essential features of the total process of continuing military readiness.

The forgetting of skills can presumably be overcome, or at least reduced in amount, by special precautions taken during training. Several possible training variations might be considered to improve skill retention, according to previously published work. Increased amount of practice, of course, is the best known suggestion for which the greatest amount of evidence exists (Naylor & Briggs, 1961; Rose, McLaughlin, & Felker, 1981). Another possibility is learning "elaboration," as suggested in the work of Rohwer (1974) and others. Still a third idea related to the maintenance of skill is a kind of "tuning" (Rumelhart & Ortony, 1977) by means of learning a schema or a "work model" (Bunderson, Gibbons, & Olsen, 1981).

Variables Affecting Skill Retention

Although each of these ways of reducing forgetting has a reasonable basis in theory, empirical findings that relate specifically to theory or to the choice of theoretical alternatives are virtually non-existent. With regard to the variable of practice, although its relation to skill retention is generally acknowledged in a qualitative sense, quantitative findings that relate the amount of practice to the amount and rate of forgetting vary widely. Although a number of studies have found a directly proportional relation between amount of practice and amount of retention (Goldberg, et al., 1981; Schendel & Hagman, 1980), a common finding indicates this relation is a decreasing one (Ammons et al., 1958; Jahnke, 1958). Mengelkoch, Adams, and Gainer (1971) found no difference in amount of retention loss for two different amounts of practice (5 and 10 trials) for several different skills of instrument flying. In contrast, the theoretical view of Wickelgren (1981) maintains that increases in the degree of learning produce large amplifications in "trace longevity," even with no difference in forgetting rate. This conception, however, refers to what Wickelgren calls "microretrieval" and so may not apply directly to the retention of skills.

A study relating retention to both amount of practice and amount of attentional effort devoted during practice was accomplished by Naylor, Briggs, and Reed (1968). The latter variable was manipulated by altering the attentional demands of a secondary task, or what the authors named "high and low coherence." The primary task was learning and retention of a moving target displayed (in three dimensions of motion) on three voltmeter dials; the secondary task was monitoring nine rows of lights by pressing buttons which turned the lights on and off according to some complex rules. The variable having the major significant effect on retention was amount of training. An interaction effect of secondary task "coherence" was indicated by the fact that retention by the 2-week training group was influenced by this factor, whereas retention by the 4-week group was not. An interesting finding was that the interval of retention (2 weeks versus 4 weeks) did not produce differential losses in retention.

A review of the Army Research Institute (ARI) research on retention of Army skills (Rose et al., 1981) identifies a number of different task characteristics that have been examined for their relation to forgetting. Perhaps most interesting of these from the standpoint of future research are (a) task learning difficulty, (b) amount of performance guidance by cues (McCluskey, Hiller, Bloom, & Whitmoush, 1978), and (c) number of task steps (Shields, Beldberg, & Dressel, 1979). These variables appear to have effects on retention in the expected direction. However, the evidence is weak in spots and specific studies of these factors would be desirable.

Other findings from a series of ARI studies are these:

1. Significant improvement in retention results from practice beyond the standard proficiency criterion of one perfect performance (Goldberg, Drillings, & Dressel, 1981; Hagman, 1980b; Schendel & Hagman, 1980).

2. Trials of practice spaced a day apart were found more effective for retention than were three trials held one after the other (Hagman, 1980a). However, a study of rifle assembly found no effects of spacing when retention was measured after 8 weeks (Schendel & Hagman, 1980).

Measurement of Skill Retention

On the whole, it is apparent that a number of important research problems remain if we are interested in identifying the variables which lead to good retention and maintenance of skills. The difficulties encountered in attempting to interpret and draw firm conclusions about skill retention appear to have several basic causes. First, empirical studies have been carried out on "skills" without agreement on a common definition of skill as a concept. Second, studies of forgetting are sometimes inconsistent in their choice of measures of amount of forgetting, rate of forgetting, and longevity. Related to this problem is the matter of choice of ways of measuring retention. In particular, measures of recall and relearning (savings) may not be as simply related to each other as suggested by the early results of Luh (1922) on nonsense syllables.

Needed Research

Drawing sensible conclusions from studies of skill forgetting is virtually impossible at the present time. The interpretation of a fundamental state of knowledge would appear to depend on the following kinds of information, but which is not available:

1. For a procedure containing four non-interfering steps, measures of accuracy of recall following 100% learning after three time intervals (immediate, 2 days, 2 weeks). Each step in the procedure should require identification of a "thing" category and an "action" category at least (e.g., turn knob A to 27°). Motor aspects of the procedure should be previously well learned.

2. Same as 1, after the time intervals immediate, 2 weeks, and 2 months.

3. Same as 1 and 2, for procedures of 4, 5, 16, and 32 steps.

4. Same as 3, with steps that are likely to exhibit interference.

These observations would simply be expected to yield some baseline data. The real problems could then be tackled. In view of this purpose, care should be taken to choose the procedural tasks carefully. Their component steps should be "realistic" ones (like setting knobs); but at the same time, peculiarities of actual job components should be avoided in the interest of generalizable results. Depending on the outcome of these exploratory studies, the following research questions can be undertaken:

1. What happens to performance of a task when practice with feedback is continued beyond the criterion of one adequate performance? What effect do differing degrees of overlearning have on retention?

2. What differential effects on retention (recall) result from task "coherence" factors, or task characteristics, that distinguish procedures from motor skills?

Skill Retention and Relearning

A number of reviews have been made of studies of retention of skills in military occupations (Hagman, Rose, McLaughlin, & Felker, 1981; Naylor & Briggs, 1961; Schendel, Shields, & Katz, 1978; Shields, Goldberg, & Dressel, 1979). It is evident that these studies have been successful in identifying a number of important variables affecting skill retention and maintenance, including level of performance attained in learning, spacing versus massing of practice, and factors in task composition.

While the amount of retention of skills, implying as it does different levels of job performance, is obviously information of value to managers of military operations, a related aspect of skill retention is at least of equal interest. This is the matter of how quickly skills can be relearned following a period of disuse. Skills acquired early in a technical school course may need to be relearned during a late period of the same course. Likewise, skills learned in a specialty school may not be required to be used immediately in a job assignment, and so may need relearning. Supervisors responsible for on-the-job training (OJT) need to be aware of the possibilities of skill deterioration and of requirements for retraining of airman job incumbents.

Most studies of skill retention assess amount retained as a recall score. A few (Ammons et al., 1958; Grimsley, 1969; Mengelkoch, Adams, & Gainer, 1971; Prophet, 1976) have additionally measured time on trials needed to retrain. Traditionally, relearning is expressed as a "saving score" (Woodworth, 1938), made into a percentage by dividing the time saved by the original learning time (or trials saved by the original learning trials). In some early studies of verbal learning, the percent of saving was found to be roughly comparable to the amount recalled expressed as a percent (Krueger, 1929). However, the saving score of relearning apparently differs greatly from task to task. For example, Ammons et al. (1958), investigating relearning of a continuous-control motor task practiced initially for 8 hours following a period of 6 months of no practice, found that initial proficiency was regained in 13 to 14 minutes. In contrast, a 15-step procedural task practiced initially for 30 trials appeared (by extrapolation) to require learners at least 15 trials to regain their original proficiency level after the same length of disuse. Expressed as percent savings, retention of the motor skill is 97%, whereas retention of the procedural skill is 50%.

Although relearning has both theoretical and practical importance, the results of existing investigations do not clearly identify the variables which determine its amount. For any given skilled performance, we are not able to predict whether it will be relearned rapidly or slowly, as compared with its original learning. From the existing evidence, it would appear that the following variables have prominent effects on relearning:

1. Level of original learning (or, amount of original practice).

2. Length of interval of disuse.

3. Composition of the skill.

In addition, the variable of amount recalled may be identified. What this means, in effect, is that the relation between amount recalled and amount of relearning required is not known. Presumably, this relationship may differ when paired with any of the variables in the previous list.

Research Problems

Besides the needed additional research on skill retention as measured by recall scores, there is a gap in knowledge to be filled by the investigation of relearning of skills. The following list represents an approximate sequence for study of these problems. Initially, it appears desirable to choose an unvarying time for a no-practice interval, in order to assess the effects of other variables. In later research, the no-practice interval may be deliberately varied in order to study its effects.

1. Relation of task composition to relearning. The composition of a procedural task can be varied in at least two ways: by variation in the number of steps involved and in the coherence of the task. Coherence is used in the study by Naylor, Briggs, and Reed (1968), where the definition is "cueing facilitation from one step to the next." In other words, a highly coherent task is one in which the cues for each next step are obvious.

a. Beginning from a level of learning, such as two performances in succession without error, compare the relearning of procedural tasks of three different lengths: 12 steps, 24 steps, and 36 steps. The no-practice interval could be 5 days.

b. With the same treatment, compare the relearning of procedural tasks of three different levels of coherence.

2. Relation of task length to relearning. Procedural tasks contain different numbers of steps. Again beginning with learning to a convenient criterion, compare the relearning of tasks of 12, 24, and 36 steps. The no-practice interval could be 5 days.

3. Relation of level of learning to relearning. Having established some baseline measures under the studies of 1 and 2, it will now be possible to select one or more learning tasks that can be used for studies relating level of learning to trials required for relearning. Although the relationship of end-of-practice performance level to amount retained (recalled) has been tested many times, it is not apparent that the same relationship will be found for relearning.

4. Relation of recall to relearning. As implied in 3, the relation of recall scores to relearning measures should be studied. It may be possible to accomplish some preliminary investigations of this relationship in connection with other studies of this sort described in 1, 2, and 3. Once the constraining factors are well understood, a direct investigation of this problem should be possible. The central question is as follows: Under what circumstance, and for what kinds of tasks, is amount of required relearning predictable from measures of recall?

5. Relation of relearning to length of no-practice interval. While virtually all of the studies described previously can be conducted with some disuse interval of a single length (such as 5 days), attention turns here to the "forgetting interval" variable itself. The various factors outlined in previous sections (task variables, learning variables) will probably be found to interact with the forgetting-interval variable, so that no simple answer to this question is likely.

REFERENCES

Benchmark Tasks for Learning Research

Foley, J. P. (1974, October). Evaluating maintenance performance: An analysis (AFHRL-TR-74-57(I), AD-A004 761). Wright-Patterson AFB, OH: Advanced Systems Division, Air Force Human Resources Laboratory.

Fox, W. L., Taylor, J. E., & Caylor, J. S. (1969). Aptitude level and the acquisition of skills and knowledges in a variety of military training tasks (Technical Report 96-6). Presidio of Monterey, CA: HumRRO Division No. 3.

Goldberg, S. L., Drillings, M., & Dressel, J. D. (1981). Mastery training: Effect on skill retention (Technical Report 513). Alexandria, VA: Army Research Institute.

Hagman, J. D. (1980). Effects of training task repetition on retention and transfer of maintenance skill (Research Report). Alexandria, VA: Army Research Institute.

Holmgren, J. E., Hilligoss, R. E., Swezey, R. W., & Eakins, R. C. (1979). Training effectiveness and retention of training extension course (TEC) instruction in the combat arms (Research Report 1208). Alexandria, VA: Army Research Institute.

Shriver, E. L., & Foley, J. P. (1974, September). Evaluating maintenance performance: The developement and tryout of criterion referenced job task performance tests for electronic maintenance (AFHRL-TR-74-57(II), Part I, AD-A004 845). Wright-Patterson AFB, OH: Advanced Systems Division, Air Force Human Resources Laboratory.

Shriver, E. L., Hayes, J. F., & Hufhand, W. R. (1975, January). Evaluating maintenance performance: Test administrator's manual and test subject's instructions for criterion referenced job task performance tests for electronic maintenance (AFHRL-TR-74-57(II), Part

II, AD-A005 785). Wright-Patterson, AFB, OH: Advanced Systems Division, Air Force Human Resources Laboratory.

Siegel, A. I., Musetti, L. L., Federman, P. J., Pfeiffer, M. G., Wiesen, J. P., DeLeo, P. J., & Shepperd, W. R. (1979, August). Criterion referenced testing: Review, evaluation, and extension (AFHRL-TR-78-71, AD-A074 539). Lowry AFB, CO: Technical Training Division, Air Force Human Resources Laboratory.

Skill Learning Beyond Initial Acquisition

Naylor, J. C., Briggs, G. E., & Reed, W. G. (1968). Task coherence, training time, and retention interval effects on skill retention. Journal of Applied Psychology, 52, 386-393.

Neves, D. M., & Anderson, J. R. (1981). Knowledge compilation: Mechanisms for the automization of cognitive skills. In J. R. Anderson (Ed.), Cognitive Skills and Their Acquisition. Hillsdale, NJ: Erlbaum.

Rumelhart, D. E., & Norman, D. A. (1978). Accretion, tuning and restructuring: Three modes of learning. In J. W. Cotton & R. Klatzky (Eds.), Semantic Factors in Cognition. Hillsdale, NJ: Erlbaum.

Shiffrin, R. M., & Dumais, S. T. (1981). The development of automatism. In J. R. Anderson (Ed.), Cognitive Skills and Their Acquisition. Hillsdale, NJ: Erlbaum.

Wickelgren, W. A. (1972). Trace resistance and the decay of long-term memory. Journal of Mathematical Psychology, 9, 418-455.

Task Difficulty as Mental Effort

Johnston, W. A., Greenberg, S. N., Fisher R. P., & Martin, D. W. (1970). Divided attention: A vehicle for monitoring memory processes. Journal of Experimental Psychology, 83, 164-171.

Johnson, W.A., & Uhl, C.N. (1976, March). The contributions of encoding effort and variability to the spacing effect on free recall. Journal of Experimental Psychology: Human Learning and Memory, 2(2), 153-160.

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice Hall.

Kerr, B. (1973). Processing demands during Mental operations. Memory and Cognition, 1, 401-412.

Moray, N. (1967). Where is capacity limited? A survey and a model. Acta Psychologica, Amsterdam, 27, 84-92.

Naylor, J. C., Briggs, G. E., & Reed, W. G. (1968). Task coherence, training time, and retention interval effects on skill retention. Journal of Applied Psychology, 52, 386-393.

Posner, M. J., & Boies, S. J. (1971). Components of attention. Psychological Review, 78, 391-408.

Rose, A. M., McLaughlin, D. H., & Felker, D. B. (1981, March). Retention of soldiering skills: Review of recent ARI research. Washington, DC: American Institutes for Research.

Shulman, H. G., Greenberg, S. N., & Martin, J. (1971). Intertask delay as a parameter of perceptual deficit in divided attention. Journal of Experimental Psychology, 88, 439-440.

Tyler, S. W., Hertel, P. T., McCallum, M. C., & Ellis, H. C. (1979). Cognitive effort and memory. Journal of Experimental Psychology: Human Learning and Memory, 5, 607-617.

Skill Maintenance and Skill Forgetting

Ammons, R. B., Farr, R. G., Bloch, E., Neumann, E., Dey, M., Marion, R., & Ammons, C. H. (1979). Long-term retention of perceptual motor skills. Journal of Experimental Psychology, 55, 318-328.

Arima, J. K., & Neil, D. E. (1978). Skill deterioration and its management. Monterey, CA: Naval Postgraduate School.

Bunderson, C. V., Gibbons, A. S., & Olsen, J. B. (1981). Work models: Beyond instructional objectives. Instructional Science, 10, 205-215.

Goldberg, S. L., Drillings, M., & Dressel, J. D. (1981). Mastery training: Effect on skill retention (Technical Report 513). Alexandria, VA: Army Research Institute.

Hagman, J. D. (1980a). Effects of training schedule and equipment variety on retention and transfer of maintenance skill. (Research Report). Alexandria, VA: Army Research Institute.

Hagman, J. D. (1980b). Effects of training task repetition on retention and transfer of maintenance skill. (Research Report). Alexandria, VA: Army Research Institute.

Jahnke, J. C. (1958). Retention in motor learning as a function of amount of practice and rest. Journal of Experimental Psychology, 55, 270-273.

Luh, C. W. (1922). The conditions of retention. Psychological Monographs, 31(142), 1-87.

McCluskey, M. R., Hiller, J. H., Bloom, R. D., & Whitmarsh, P. J. (1978). Skill decay of sixteen common tasks for MOS 11B and 11C (Final Report). Alexandria, VA: Army Research Institute.

Mengelkoch, R. F., Adams, J. A., & Gainer, C. A. (1971). The forgetting of instrument flying skills as a function of the level of initial proficiency (NAVTRADEVCEN-71-16-18). Port Washington, NY: Naval Training Device Center.

Naylor, J. C., & Briggs, G. E. (1961). Long-term retention of learned skills: A review of the literature (ASD Technical Report 61-390, AD-267 043). Wright-Patterson AFB, OH: Aerospace Medical Research Laboratory.

Prophet, W. W. (1976). Long-term retention of flying skills: A review of the literature. (Final Report 76-35). Washington, DC: Human Resources Research Organization.

Rohwer, W. D., Jr. (1974). Elaboration and learning in childhood and adolescence. In H. W. Reese (Ed.), Advances in child development and behavior, (Vol 8). New York: Academic Press.

Rose, A. M., McLaughlin, D. H., & Felker, D. B. (1981). Retention of soldiering skills: Review of recent ARI research. Washington, DC: American Institutes for Research.

Rumelhart, D. E., & Ortony, A. (1977). The representation of knowledge in memory. In R. C. Anderson, R. Spiro, & W. E. Montague (Eds.), Schooling and the Acquisition of Knowledge. Hillsdale, NJ: Erlbaum.

Schendel, J. D., & Hagman, J. D. (1980). On sustaining procedural skills over prolonged retention intervals (Technical Report). Alexandria, VA: Army Research Institute.

Shields, J. L., Goldberg, S. L., & Dressel, J. D. (1979). Retention of basic soldiering skills (Research Report 1225). Alexandria, VA: Army Research Institute.

Wickelgren, W. A. (1981). Human learning and memory. Annual Review of Psychology, 32, 21-52.

Skill Retention and Relearning

Ammons, R. B., Farr, R. G., Block, E., Neumann, E., Dey, M., Marion, R., & Ammons, C. H. (1958). Long-term retention of perceptual-motor skills. Journal of Experimental Psychology, 55, 318-328.

Grimsley, D. L. (1969). Acquisition, retention, and retraining: Training Category IV personnel with low fidelity devices (Technical Report 69-12). Presidio of Monterey, CA: HumRRO Division No. 3.

Hagman, J. D., Rose, A., McLaughlin, D. H., & Felker, D. B. (1981). Retention of military tasks: A review. Alexandria, VA: Army Research Institute.

Krueger, W. C. F. (1929). The effect of overlearning on retention. Journal of Experimental Psychology, 12, 71-73.

Mengelkoch, R. F., Adams, J. A., & Gainer, C. A. (1971). The forgetting of instrument flying skills. Human Factors, 13, 397-405.

Naylor, J. C., & Briggs, G. E. (1961). Long-term retention of learned skills: A review of the literature (ASD Technical Report 61-390, AD-267 043). Wright-Patterson AFB, OH: Aerospace Medical Research Laboratory.

Naylor, J. C., Briggs, G. E., & Reed, W. G. (1968). Task coherence, training time, and retention interval effects on skill retention. Journal of Applied Psychology, 52, 386-393.

Prophet, W. W. (1976). Long-term retention of flying skills: A review of the literature (Technical Report 76-35). Alexandria, VA: Human Resources Research Organization.

Schendel, J. D., Shields, J. L., & Katz, M. S. (1978). Retention of motor skills: Review (Technical Paper 313). Alexandria, VA: Army Research Institute.

Shields, J. L., Goldberg, S. L., & Dressel, J. D. (1979). Retention of basic soldiering skills (Research Report 1225). Alexandria, VA: Army Research Institute.

Woodworth, R. S. (1938). Experimental psychology. New York: Holt.

CHAPTER 4

TRAINING EXPERT APPRENTICES

Robert Glaser
Learning Research and Development Center
University of Pittsburgh

What are some fundamental considerations in the design of an Air Force laboratory for research in learning and training?

I believe that the laboratory should focus on the question of how to produce expert apprentices. By "expert apprentices" I mean people who, when they leave training, have strong capabilities to profit from the experience they receive in their military careers. What kind of basic and more advanced knowledge and skills, what kind of learning skills, what perception of their jobs should they be given in school so that they can profit maximally from later experience?

There has always been a debate about the relationship between what students learn in school and what they need for the job; school is just school, and experience is necessary. So it seems to me the mission of school should be to produce people who are highly capable of profiting from work in the field. That is what I mean by producing expert apprentices; i.e., novices who have an adequate information base and skills for learning new things. Apprentices can do many things, but they are not experts; they are far from experts, and on-the-job training and experience can take place efficiently with appropriately trained apprentices.

A key objective of schooling is to train people in abilities that enable them to profit from the experience after school. The differences in the knowledge structures and perceptions of novices and people who have some degree of experience are so drastically different that we have to put our apprentices in a position where they can acquire most readily the kind of teaching that comes from experience. Apprentices need to learn as they work in their field units. We want to maximize this benefit through what is taught in initial training.

In order to think through this problem, we first have to ask what it is that is acquired by experience, and then we have to see whether it is possible to do things in training that maximize what students are supposed to get out of experience. That is the line of this logic. First, I discuss what is acquired by experience and then list suggestions for research.

What Is Acquired Through Experience?

Categorization Ability

People who are experienced are very good categorizers; that is, when meeting up with a situation, they encode it into a category of knowledge that has included in it ways of handling or solving the problem. Experts are excellent categorizers, as shown in the expert-novice chess studies. These experts do not have super processing powers. They recognize a configuration of pieces, and that configuration has meaning - like a word has meaning - and they know what to do with it. Through experience, they become expert at recognizing and categorizing things; and there is a great deal of automaticity to this performance. For example, an experienced jet engine mechanic rapidly categorizes a reported engine malfunction into a class of repairs that has certain characteristics and not others (e.g., core engine damage requiring 1 week of in-shop repair); an experienced manager classifies a loss of morale as requiring departmental reorganization rather than personnel dismissal. In both of these cases, features of the problem situation result in the accessing of categories in memory that are attached to probable solution procedures.

Principled Knowledge

Categorization has certain additional properties. It involves sorting on the basis of "principled" knowledge versus what can be called "literal" knowledge. In the physics studies, for example, when a novice is asked to look at a problem, he or she says it is a pulley problem, or it is an inclined plane problem; but when you ask an experienced individual to look at the problem, he or she says it is a Newton's Second Law problem, or it is a force problem. There is a relatively automatic categorization of the problem, but it is based on principled knowledge rather than literal feature knowledge. This is the kind of proficiency that should be encouraged by experience. The question is: What basis for it might be supplied in training? This is the problem I am trying to pose. Principled knowledge is acquired both in school training and as a result of on-the-job experience. However, school training can provide a basis that fosters or inhibits its development; e.g., the kind of understanding of electronic principles that an individual acquires in training will influence the way in which experience with on-the-job troubleshooting problems is organized for further use.

Mental Models

Another thing that people gain from experience is adequate theories or mental models of their work; they have especially effective ways of representing the situations that confront them. Highly proficient people have models that work, schemata that are appropriate; and these working theories are more complete than those of the less experienced. These theories or models that people have in their minds are used as scientists use theories when they think about a problem. They use a theory to make predictions, to think with, and to say whether something else should be present in the situation that is not apparent. A theory of a task works like a cognitive schema; it is a structure of knowledge that is used to think with and to remember things. Schemata or models (or organizations of knowledge) are constantly amended through on-the-job experience. Models of knowledge are challenged in the course of experience. If the model does not conform to real situations, changes are made or reality is assimilated to the model. When novices come into a situation, they too have mental models - naive ones which are relatively incomplete and inadequate - but it is these theories or knowledge organizations that have to be addressed and developed through appropriate teaching.

Automatization

Experts display a good deal of very rapid performance without apparent cognitive effort. They "parallel process" such that various attentional requirements do not interfere with one another; highly skilled components of performance occur with little apparent conscious processing. In particular, the basic skills of the expert are automatized, such that mental space is left over for aspects of performance that require conscious processing. (Like tennis professionals, they have their basic strokes well practiced so they do not have to worry about their backhand and can concentrate on the strategy of the game.) Another dimension of training is to assist in the development of automaticity in basic skills so that the time left over for higher-level job performance is maximized.

Proceduralization

The knowledge of an expert is proceduralized. Experts not only know things, but they know when that knowledge is applicable; they know when it is to be used and when it is not to be used. The expert's knowledge is embedded in procedures for the use of that knowledge. Expert solvers of physics problems not only can identify a Newton's Second Law problem, but they can also can indicate the conditions under which a form of Newton's Second Law is applicable. Their knowledge is proceduralized by being attached to a set of conditions for its use. (If you think in terms of production system statements such as "if-then," then you know that the expert's knowledge not only has the "then," but also has the "if" attached to it. The novice's performance is less active. It is largely "then." There is not much "if" related to it.) If the conditions are such and such, then you do that; this is what I mean by proceduralized knowledge. The training question then becomes: How do we produce knowledge in novices so that it becomes proceduralized?

Metacognition and Learning Skill

Another thing that is important to acquire in training is the capability to learn from experience. We have to understand what this capability is. To some extent, it consists of a capability for monitoring one's own learning, or of interrogating learning, so that the experience obtained can be managed and assimilated. It is this kind of metacognitive skill that good learners display. When they hear something, they check on it, they try to reconcile it with the job schemata they have, and they check on the products of their problem solving. Mature learners acquire these capabilities, and an objective of training is to help novices begin to develop the same capabilities.

Training

The above performance factors are what experts learn after time on the job, and the objective of training is to produce people who profit from job experience so that they develop these characteristics of expertise. Now comes the crucial task: What do you do in the formal training in the military to achieve this?

There are cycles of training, first on-the-job and then back to retraining, but a stage can be defined as training and then on-the-job training. Therefore, the research issue becomes how to design training and research on training (within a stage) that addresses these objectives of developing expertise. How can training encourage the capability to profit from experience that yields high levels of proficiency? One clue has already been given to us by what Ray Christal said. He is thinking about assessing automatization or the progress in automatization. In this regard, the work by Wally Schneider (1982) on automatization is very interesting. He indicated that the way to produce "unconscious processing" is to find some consistency of performance, because inconsistency requires conscious processing. Schneider studies how air traffic controllers determine intercept paths. The tasks look varied, but he found a consistent principle that these people can use and that can be repetitively practiced under speed stress, so that it becomes highly automatized.

The idea is to determine the conditions that produce automaticity and ways of assessing it to insure its reliable presence. It is possible that when basic skills are measured when people leave school, a proficiency examination should assess the degree to which some of their knowledge or skills have been automatized. For example, long-term studies of reading show that students who develop automatized skills in the early grades may be those who in the long run develop comprehension skills faster. This suggests that during training, assessments could be made of the level of automatization characteristic of those students who develop better and more complex knowledge later in training or on the job.

Problem Solving and Knowledge

Learning skills and problem-solving skills can be developed and strengthened in the course of training. In the course of acquiring information and skill, students can interrogate a body of knowledge, solve problems about the application of their knowledge, make up theories and extrapolate from the information they have. Students have models of what they are doing, and the opportunity to interact with the model develops the capability for interrogating their own knowledge in a way that leads to learning.

There currently are a number of widely used training programs based on information-processing psychology and older psychometric theory that attempt to train people in general problem-solving heuristics. However, the theory on which these programs is based is changing. New theory emphasizes the importance of learning problem-solving as a person acquires knowledge. For example, in the artificial intelligence field, there has been a shift from building machines with big search algorithms that could search faster, with quicker hardware, to the building of better knowledge structures that can be readily accessed. Knowledge and process are highly interactive in human performance.

Knowledge structure interacts with memory skills. For example, in a study of chess knowledge, the investigator took children who knew chess well and adults who did not, and gave them two memory tasks: (a) remembering a briefly presented chess board and (b) performing a digit span task of the kind that appears on standardized intelligence tests. The results showed that the children had very good memories, and the adults had poor memories, for chess positions. On the digit span test, however, the children had much more limited memory than did the adults (the usual test finding on the knowledge-lean digit task). Therefore, there are interactive effects of knowledge and process that have to be considered in training. Knowledge is not just inert, waiting to be used; rather, knowledge is an active shaper of cognitive process.

These are very complex problems, because it is obvious that some people are very good at employing generalized processes when they confront a situation where they do not have any or much knowledge. However, when they confront a situation where they have knowledge that can be brought to bear, they do not use these general heuristics; rather, they employ knowledge-specific processes that are more powerful in particular situations than the more general methods. Thus, problem solving and proficiency in higher-level processes are best acquired in the context of a knowledge base. The kind of instruction that this implies is not unlike the kind being thought about in intelligent computer systems, w ere people can constantly interrogate and inquire in order to learn.

There is a good article by Collins and Stevens (1982) that analyzes the method of inquiry teachers. What do inquiry teachers do? They force the student to generalize; they give counterexamples; they reduce an argument to a Reductio Ad Absurdum to show it is not a complete case. Inquiry instructors use entrapment strategies. They try to make explicit the incomplete knowledge of students by forcing them to apply it in carefully selective situations. This form of interactive instruction develops students' attitudes toward interrogating their knowledge. They learn to reason and to come up with theories about what they think they know. They ask the instructor to interrogate them or interrogate their knowledge base themselves so that their ideas get changed. In this way, students develop a capability for flexibility in which their knowledge can be amended in one way or another in the course of training and job experience.

Naive and Pedagogical Theories

At the beginning of instruction, students have some kind of a beginning knowledge structure and a theory of this knowledge that has to be taken into consideration. One way to do this is to design "pedagogical" theories that facilitate learning. The instructor gives the student a theory that is easy to manage - a theory that can be used to remember and access knowledge, to exercise knowledge, to work with knowledge - but which is not necessarily an accurate theory. (It is a pedagogical theory that will soon be overthown.) Instructors can present such theories to a class or to an individual. Good teachers do this all the time. They say, for example, "Let us assume that what we are learning operates this way." This tactic gives students a structure to think with, but the teacher knows full well that it is only an interim pedagogical theory. It seems to me that teachers (rather than saying, as many teachers do, when a student has trouble in learning, "Let me give you an example") should say "Let me give you a theory." They should say, "Let me give you a structure in which you can put what you know." I do not mean to be too sophisticated when I use the word "theory." For example, people have a model or theory of the task for a series of procedures when they operate a piece of equipment.

At the beginning of instruction, people have theories that are very persistent. There are studies on "naive physics" theories that students believe. These investigations took college and high school students and described situations to them (e.g., a steel ball in a tube going around in a spiral) and asked them questions (e.g., what direction the ball would continue to go when it emerged from the tube). A variety of situations like this were presented. The answers the students gave were inadequate Aristotelian theories of motion (which are intuitively what people think should happen), but were not what really happens in Newtonian mechanics. The actual theories of physics are very nonintuitive, and students come to instruction with naive theories that a teacher needs to understand in order for instruction to proceed properly. Studies have been carried out on the kind of theories of electronic circuits with which students come to electronics training.

In physics it is possible for students to come from a course with a whole new set of words and some ability to solve physics equations, but still believing in the old theory. The significant question for training research is how to produce novices who have reasonably useful theories that they can use to learn from experience. Theories can be taught as entities of explanation, along with a description of the operation of a system. Theories can also be taught as models that can be questioned and tested against certain obtained information. This latter form of instruction encourages new experiences to be assimilated into integrated knowledge structure. Theory developed in this way results in better memory storage and accessibility of practical procedures associated with particular theories of operation. (See the earlier Problem Solving and Knowledge section of this paper and the research suggestions section at the end.)

Theory and Practice in Training

One has to look very carefully at the various occupational specialties in the Air Force and begin to ask whether theory-based instruction is useful for best understanding and performance of a job. We need to concentrate on the development of rapid automatization and less conscious processing in certain areas of a job. For others, there are the kinds of decisions that take thinking time, so there should be a balance in certain job situations between teaching in relation to theories and models and teaching in a form of drill and practice. This is complicated because even simple operating procedures are often best performed if one has a theory or model in mind. There are studies that show that procedures are better carried out when one begins with some kind of advance organizer rather than immediately with the first step. In an assembly task, rather than doing step one, step two, etc., it may be better to start with the idea of where the assembly is to end up, so that some kind of theory guides performance.

Thus, different job specialties have various dimensions, and training questions arise. For example, is it worthwhile to use interactive model-driven forms of instruction, or is an interactive form of instruction too time-consuming for certain jobs? This is a matter for training research.

It is useful for training research to consider the conflict between theory and automatization. For example, in learning mathematics there are conflicting demands for mathematical understanding and computational skill. Recent research is making clear that to compute with numbers (in word problems for example) one has to have certain forms of understanding and that the debate between computational skill and understanding is not a useful discussion to continue. In mathematics and other subject matters, understanding and skill are mutually supportive and are not an all-or-none matter. Understanding, such as knowledge of the function of place value in numbers, makes one better in subtraction, encourages faster learning and results in fewer bugs in the course of learning.

Understanding and learning are related to "naive theory" which has been mentioned above. "Naive theory" can be defined as the intelligent systematicity of the beginning learner. This is the basis for "buggy analysis" by some researchers. Based on this work, an important kind of research for students coming into the Air Force in a certain training program is to make a collection of the systematicities of thinking that these students bring to training. What kinds of bugs do they generate in the course of learning? A catalog should be made of these initial conceptions so that these ideas can be treated as forms of thinking and not as stupid, lazy errors. Students are intelligent people and have forms of thinking that need completion. One example is in writing, where students have a conflict between their naive spoken language and the formal demands of written language. In the course of this they develop a kind of metalanguage that is between the two. This results in forms of writing prose that are forms of one's speech and are intelligent ways of handling oral language, but which are not intelligent ways to handle the formal writing system. If these are understood as systematic errors, the instructor can work with systematic ways of interrogating the students' conceptions. The instructor's attitude can be that the student has a form of knowledge or a theory that is incorrect and that can be repaired through appropriate training.

In general, problems in training should be driven by the objective of developing intelligent apprentices. The objectives involved are those described above. These dimensions of expertise should influence instruction and research on training.

Macroprocess in Training (School and Classroom Variables)

So far we have been considering what can be called the microprocesses of learning or cognition as these relate to proficiency in performance. There is also a kind of work in training research that involves the macroprocesses of instruction. The macroprocesses of instruction are the teaching and management variables of the school and classroom. These include variables such as: time on task, interaction between student and teacher, and form of examination. While the macroprocess of teaching and the microprocess of learning interact, much progress can be made by operating on the macro level. Just as the steam engine was built without a theory of thermodynamics, we need to become more effective in work at the macro level. There is a long history of studies of good and bad teaching that are not very informative, but this field is becoming more sophisticated and is beginning to look at classroom processes in more potentially useful ways. For example, a recent research project has analyzed video-tape records of good and poor teachers, in terms of the planning structure methodology of artificial intelligence. In these tapes, you see teachers conducting a lesson in mathematics and can examine the details of student interaction and lesson management. The planning structure of the good teacher is analyzed in detail and compared with the planning structure of the less experienced teacher. Certain activities become apparent, such as subtle forms of individualizing and adapting to student requirements in the context of group instruction. These data would be very useful for instructor training.

These macrovariables of teaching such as classroom management, the design of the motivational environment, peer help in the classroom and the other such factors, if attended to in research, might well result in a better understanding of the dynamics of the task of the instructor in the classroom. Thus, I would advocate research on both the macrovariables just described and cognitive microvariables previously mentioned.

Research and Development

How should scientists relate to development? Please remember that knowledge in the field of behavioral science, for the most part, is not placed in rigorous enough form that it can be transmitted in a way which is not subject to possible misinterpretation in development. There are subtleties and ambiguities involved. A research psychologist can say that I know something about the acquisition of vocabulary and text comprehension, and that as a result of this information it seems to be a good idea to design instruction in this way. When someone tries to pick this up, they frequently caricature the researcher's idea. There was some documentation on this matter in the early years of the space program when it was found that if the scientist who developed an original idea participated in some way through the stages of development, then these stages were carried through with more efficiency and success. At least the opportunity for the researcher to observe the implementations of the original ideas appears to be an important aspect of the research and development continuum.

Research Suggestions

In this section, I list research ideas related to sections of this paper concerned with what is acquired through training and the conduct of the training process. Research questions or topics are briefly listed, and reports of relevant research are given that illustrate how questions might be formulated and indicate methodologies of approach that might be used. The reports referred to are in the reference list that follows.

Categorization ability. How is knowledge categorized and how do these categories change as a result of training and experience? What is the difference between novices and experts, and how might knowledge of this difference indicate how training should proceed? Representative articles include Chi, Feltovich, and Glaser (1981); Chi, Glaser, and Rees (1982); Egan and Schwartz (1979); and Paige and Simon (1966).

Principled knowledge. In addition to the procedures required for performing a task, what perceptions and understanding of the task must be learned to ensure that these procedures can be carried out rapidly with a minimum of error? A representative article by Riley, Greeno, and Heller (1983) examines children's problem-solving ability, but the methodology could be very useful to consider for technical tasks.

Mental models. It is important to determine the kinds of mental structures that influence performance. How can these be conceptualized and identified? A representative book by Gentner and Stephens (1983) covers a wide variety of studies including mental models in understanding electricity and physical systems.

Automatization. What conditions of instruction favor the development of automaticity in basic skills? How is this related to the development of higher levels of job performance? Representative studies include Perfetti and Lesgold (1979), on the interaction of basic and higher-level skills and verbal comprehension; Schneider (1982), on training for automaticity; and Shiffrin and Schneider (1977), on general theory.

Proceduralization. Experts develop not only declarative knowledge but declarative knowledge that is attached to conditions of its use so that knowledge becomes effectively useable. Representative studies include Anderson (1982), an article on theory; Chi, Glaser, and Rees (1982); and Lesgold (in press).

Metacognition and learning skill. How are self-regulatory and self-management skills that influence learning and problem solving influenced through training? Representative studies include Brown (1978), on theory and methodology; Brown and Palincsar (1982); and Gitomer and Glaser (1983).

Training. In this section of the present paper, I have concentrated on automaticity, and the references given under the heading of automatization are relevant here.

Problem solving and knowledge. How is knowledge acquired so that it is effectively used as a basis of learning and problem solving? Representative studies include Collins and Stevens (1982); and Glaser (1983).

Naive and pedagogical theories. How can the naive theories of a learner be determined, and how can they be used as a basis for instruction? A representative article is by McCloskey, Caramazza, and Green (1980).

Theory and practice in training. In this section, I refer to the importance of determining the intelligent "bugs" of the learner that can be used as a basis for training. Representative articles include Brown and Burton (1978); and Glaser (1981). Also in this section, I refer to the macrovariables of teaching and classroom management. A representative article by Glaser (1982) has a section on school processes with a variety of references.

REFERENCES

Anderson, J. R. (1982). Acquisition of cognitive skill. Psychological Review, 89, 369-406.

Brown, A. (1978). Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), Advances in instructional psychology (Vol. 1). Hillsdale, NJ: Erlbaum.

Brown, A., & Palincsar, A. S. (1982). Inducing strategic learning from texts by means of informal, self-control training. Topics in Learning and Learning Disabilities, 2, 1-17.

Brown, J. S., & Burton, R. R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. Cognitive Science, 2, 155-192.

Chi, M. T. H, Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. Cognitive Science, 5, 121-152.

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), Advances in the psychology of human intelligence (Vol. 1). Hillsdale, NJ: Erlbaum.

Collins, A., & Stevens, A. L. (1982). Goals and strategies of inquiry teachers. In R. Glaser (Ed.), Advances in instructional psychology (Vol. 2). Hillsdale, NJ: Erlbaum.

Egan, D., & Schwartz, B. (1979). Chunking in recall of symbolic drawings. Memory and Cognition, 7, 149-158.

Gentner, D., & Stephens, A. L. (1983). Mental models. Hillsdale, NJ: Erlbaum.

Gitomer, D., & Glaser, R. (1983, May). Mental models, self-regulation, and instruction. Paper presented at the Conference on Aptitude, Learning, and Instruction: Conative and Affective Process Analysis, Stanford.

Glaser, R. (1981). The future of testing. American Psychologist, 36, 923-936.

Glaser, R. (1982). Instructional psychology. American Psychologist, 37, 292-305.

Glaser, R. (1983, June). Education and thinking: The role of knowledge (Technical Report PDS-6). Pittsburgh: University of Pittsburgh, Learning Research and Development Center.

Lesgold, A. M. (in press). Acquiring expertise. In J. R. Anderson & S. M. Kosslyn (Eds.), Tutorials in learning and memory: Essays in honor of Gordon Bower. San Francisco: W. H. Freeman.

McCloskey, M., Caramazza, A., & Green, B. (1980). Curvilinear motion in the absence of external forces: Naive beliefs about the motion of objects. Science, 210, 1139-1141.

Paige, J. M., & Simon, H. A. (1966). Cognitive processes in solving algebra word problems. In B. Kleinmutz (Ed.), Problem solving. New York: Wiley.

Perfetti, C. A., & Lesgold, A. M. (1979). Coding and comprehension in skilled reading. In L. B. Resnick & P. Weaver (Eds.), Theory and practice of early reading. Hillsdale, NJ: Erlbaum.

Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. In H. P. Ginsburg (Ed.), The development of mathematical thinking. New York: Academic Press.

Schneider, W. (1982, April). Automatic/control processing concepts and their implications for the training of skills (Technical Report HARL-ONR-8101). Champaign, IL: University of Illinois, Department of Psychology.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.

CHAPTER 5

CONTRIBUTIONS OF COGNITIVE PSYCHOLOGY TO LEARNING FROM INSTRUCTION IN AIR FORCE TRAINING

Thomas J. Shuell
State University of New York at Buffalo

During the past 10 or 15 years, there has been a tremendous shift in the way psychologists think about issues related to learning and training. This change in perspective has been heavily influenced by two separate but related revolutions: the computer revolution with its emphasis on information processing and the cognitive revolution with its emphasis on mental activities. The impacts of both have been widespread. On topics closely associated with training, for example, the influence can be seen clearly in current thinking about the nature of mental abilities and psychometric tests, as well as in current thinking about the nature of human learning.

This paper will focus primarily on the major themes evident in modern-day cognitive psychology and the influence they have had on the way we currently think about human learning and instruction. At present, for example, learning is generally considered to be an active, constructive process rather than a passive, reproductive process (as was the general case 10 to 20 years ago). The emphasis is no longer strictly on behavior; rather, it is on the mental processes and knowledge structures that can be inferred from behavioral indices. This change in thinking about the nature of learning is beginning to produce corresponding changes in the way we think about instruction and training. There is increasing emphasis, for example, on the way specific content and instructional procedures engage and elicit the psychological processes and knowledge structures appropriate for the desired learning outcome to be achieved.

These various changes in the way psychologists think about learning, and its relationship to instruction, need to be considered in establishing a research agenda for a learning laboratory of the type being discussed here. The purpose of research conducted in such a facility is to provide a theoretical and empirical knowledge base for improving the effectiveness of Air Force training, both in the more formal schools and in the less structured training that occurs "on the job." If we are to ask the best and most appropriate research questions, then we need to have some idea of the various factors that influence learning in the different training environments found in the Air Force, how these factors are interrelated, and how it might be possible to integrate them in a meaningful way. The next section of this paper will discuss some of these general issues in more detail. Subsequent sections will discuss specific ways in which cognitive psychology has influenced current research on learning and some of the specific research issues that might be considered in an Air Force learning research laboratory.

Learning from Instruction

Two characteristics of traditional research on learning and cognition - especially the better, more rigorous scientific research - seriously limit the usefulness of this research when it comes to solving practical problems such as those involved in trying to improve the effectiveness of Air Force training. The first of these is the lack of an explicit concern for instruction and application. The second is that most of this research (as is the case with most basic, scientific research) is descriptive in nature (i.e., the goal of this research is to identify and describe the various factors that influence the learning process, with very little, if any, thought given toward how the findings of this research relate to the issues involved in using it to improve training outcomes). That is, the data and conclusions produced by this traditional type of research are not completely appropriate for the task at hand. At times, it is almost as if we have a good answer, but the answer is to a different question than the one in which we are really most interested.

Most individuals concerned with applying scientific knowledge to the solution of practical problems have envisioned a linear model in which basic research findings move through a series of stages or steps toward ultimate implementation and dissemination in a more-or-less linear sequence (e.g., Hilgard, 1964). I have become thoroughly convinced, however, that this linear model is totally inappropriate for any kind of meaningful research or development with regard to practical problems (Glaser, 1973; Shuell, 1982). We need to stand the whole research and development process on its head and realize that basic research in psychology is a descriptive science whereas what we need for improving training and education is a prescriptive science. The nature of a prescriptive science is very different from a descriptive science, and these differences should be taken into account when planning for research on learning from instruction (Shuell, 1981b, 1982).

Descriptive vs. Prescriptive Science

Science, as we normally think of it, is concerned with understanding the nature of naturally occurring phenomena. This understanding typically takes the form of an accurate (i.e., reliable and valid) description of the phenomenon being investigated. Typically, it is assumed that there is a single best answer or description of the phenomenon; all we have to do is discover it. For purposes of a descriptive science, for example, it is sufficient to show that the way in which an individual represents a problem determines how successful he or she will be in finding a successful solution to the problem. While various issues, such as the conditions necessary for obtaining the effect, the extent to which various factors may influence the likelihood of obtaining the effect, and so forth, may be investigated, the purpose of the research is to describe the relationship by determining the extent and conditions under which the phenomenon occurs.

For instructional purposes, however, this description is inadequate. Instead, what we need is empirical evidence on the best way for a student to represent a problem in a given situation and on what techniques or methods an instructor could use to help him or her represent the problem in the most appropriate and efficient manner. A prescriptive science, in contrast to a descriptive science, recognizes that there can be more than one scientifically valid way of accomplishing the same end and seeks to provide theoretical models and empirical data directly relevant to the task faced by the individual who must make decisions about what to do in a particular situation. For the most part, an integrated and systematic body of scientific knowledge appropriate for this type of practical decision making simply is not available at the present time.

The chief purpose of a prescriptive science is to create artifacts that will serve some purpose (Simon, 1981). That is, things that do not exist naturally in nature are created or invented in order to accomplish something we want done. For example, a water wheel is built (nothing occurs in quite that form in nature) in order to turn a grain mill or an electric generator. Similarly, an instructor may decide to lecture on a particular topic in order to have trainees learn certain information. Usually there is more than one way in which such a purpose can be accomplished; the grain mill can also be turned by oxen walking in a circle, and the information to be learned in class may be equally well acquired by having the trainees work on a mockup or by discussing a specified problem in small groups.

While some methods for accomplishing a particular purpose may be better than others, frequently the criteria involved in identifying the "better" or "best" approach are complex, with a variety of tradeoffs being involved. Often the only thing that can be said is that the various approaches represent different ways of going about accomplishing the task, each with different consequences and side effects (Shuell, 1982). Research on human learning with the overall purpose of ultimately improving the effectiveness of training or instruction should be designed with these various factors explicitly in mind. One way to accomplish this goal is to develop a conceptual model that can assist in guiding both basic and more applied research on learning from instruction.

Toward a Prescriptive Model of Learning from Instruction

Recently, I have developed a model that incorporates these considerations (Shuell, 1981b). The model is based on the notion that in order to make useful instructional decisions based on valid scientific findings, we need to develop a knowledge base that recognizes from the very beginning that there are two types of constraints or "givens" whenever one makes an instructional decision.

The first constraint or given specified by the model is the nature of the outcome. What exactly are we trying to help the students achieve? Is it a cognitive goal, an affective goal, the removal of a perceptual problem or misconception, or something else? Various types of knowledge characteristic of particular instructional outcomes can be identified, and some of these different types of knowledge will be discussed in more detail later in this paper. The concern for the nature of the desired outcome extends beyond the mere specification of instructional objectives. The model emphasizes the importance of identifying the nature of the instructional goal in a psychological manner that is consistent with (a) the other given specified by the model (to be discussed next) and (b) the selection of an instructional intervention that appropriately matches the two givens and that has a high probability of leading to the efficient attainment of the goal. It is likely that many or most of these outcomes can come from existing analyses and descriptions of Air Force jobs and job composites, (Gagné's paper previously discussed these types of outcomes).

The second given identified by the model has to do with the nature of the existing situation. When making decisions about the most appropriate type of training to provide, one is faced with the realities of the situation in which the training must be accomplished, and these realities place constraints on the type of instruction that can provide the most effective training, given that situation. Included in this component of the model are many more considerations than the traditional concern for the prior knowledge of the learner. These additional factors include concerns for the resources one has available, concerns for the social dynamics that exist among the trainees and between the trainees and the intructor, and so forth.

Now, given these two types of constraints, what knowledge about the psychology of learning from instruction does an individual need to have in order to make a viable instructional decision based on the best scientific evidence available? When faced with making a decision about training, one might ask: What am I going to teach next; and given that I am going to teach this material or skill next, how am I going to teach it in order for learning to be most effective? In making these instructional decisions, one is faced with teaching particular outcomes within an existing situation, and those making the decision need to know something about alternative procedures that will help the students to get from Point A to Point B. Thus, a means-end analysis needs to be performed, and a viable theory of instruction - or systematic body of psychological knowledge - must be able to provide scientific data relevant to this task.

Unfortunately, most of the scientific knowledge currently available is not organized in a manner that lends itself to making these types of practical decisions. What is needed is research that can say: Given these outcomes, .d given this situation, these limited number of possibilities exist for effective instruction. From that small pool of alternative procedures, a person can then make a reasonable professional judgment. We probably will never reach the point where we are capable of prescribing specific instructional procedures, but I do think there is a possibility of developing a body of knowledge that will help us to narrow the alternatives in a scientifically valid manner. In order to accomplish this goal, however, we must set out to do research conceptualized in a manner appropriate for the task.

In discussing research on learning from instruction, some mention must be made of the findings generated from classroom research, especially the process-product research (e.g., Brophy, 1979; Brophy & Good, in press). Until recently, classroom research and learning research have been isolated from one another. Although virtually none of the classroom research has been conducted in a context similar to the one encountered in Air Force training, many of the findings can undoubtedly be generalized in some way to Air Force training environments. In any case, truly productive research on learning from instruction must be performed with an awareness of the research being performed on classroom learning, and research performed at the Air Force learning research laboratory should reflect this awareness.

Contributions of Cognitive Psychology

Cognitive psychology has brought a number of identifiable concerns or themes to the field of psychology in general, and educational psychology in specific. These concerns have emphasized new aspects of human information processing and learning that need to be incorporated into any attempt to understand the nature of effective training and learning from instruction. Some of the topics emphasized by modern-day cognitive psychology include a concern for understanding what one is learning (i.e., comprehension), knowledge representation, problem solving, and the nature of individual differences in learning from instruction. These topics will be discussed next, and an attempt will be made to relate the concerns represented by these topics to research that could be conducted in the proposed laboratory.

Comprehension

Probably the first and most important of the issues emphasized by cognitive psychology is a concern for meaning - that is, the manner in which an individual extracts meaning from some experience. The emphasis here is on understanding, rather than merely on learning (or learning how to perform) the material or task being acquired. This emphasis includes a concern for comprehension - specifically, or at least most extensively, for prose comprehension.

There has been a great deal of research on written comprehension (e.g., Reder, 1980; Spiro, Bruce, & Brewer, 1980). Very little work, however, has been done on oral comprehension, which is equally important, especially since we convey as much of our instruction via the oral modality (i.e., lectures, etc.) as we do by the written modality (i.e., textbook, handouts, etc.). A few studies (e.g., Anderson & Biddle, 1975; Smiley, Oakley, Worthen, Campione, & Brown, 1977) have compared comprehension in the two modalities without finding any large differences. Researchers in this area frequently make the assumption that the same psychological mechanisms are responsible for both written and oral comprehension. While this assumption does not seem totally unreasonable, it is an issue that we really have not pursued very far. Given its importance for training and learning from instruction in a variety of situations (including on-the-job training), I think it is worth further investigation. That is, what is the nature of oral comprehension, as well as the nature of written comprehension; and what is the role of each in learning from instruction in a training environment?

A third aspect to this question of comprehension is a concern for visual comprehension. This modality is one about which we know very little with regard to what constitutes comprehension, outside of the traditional work on spatial abilities, which is not extremely relevant to the concerns being discussed here. There are some very interesting indications, however, that visual comprehension may be a useful topic to represent knowledge for instructional purposes, which is the larger concern in this context. It may be, for example, that in some cases we could represent the knowledge that a trainee needs to acquire in order to learn specific skills in spatial formats that would lead to more rapid acquisition than would be the case if we presented the instruction in the written or oral modalities. There is some evidence (e.g., Levin, 1981) which shows that visual mnemonics can be very powerful instructional tools. There is also work by Lynn Cooper (1982) at the Learning Research and Development Center (LRDC) at the University of Pittsburgh which identifies different types of visual processing strategies that individuals use and some indication of what occurs when there is a match or mismatch with regard to instructional presentations - that is, what happens if you present instruction that is consistent with an individual's preferred perceptual processing strategy and what happens if you do not.

From an instructional point of view, there has been a fair amount of research on aids to comprehension - things like advanced organizers, adjunct questions, and so forth. Rich Mayer (in press) recently has completed a review of the literature on comprehension aids, and a recent paper by Merl Wittrock (1983) discusses many of the same issues from a slightly different perspective. In addition, the visual mnemonics discussed in the preceding paragraph can also be regarded as one type of comprehension aid. The concern for which instructional aids might facilitate learning and comprehension in the types of training environments encountered in the Air Force is something that should be looked at in terms of the various modalities of comprehension discussed above, for such aids are an instructional tool that deserves further investigation.

Types of Knowledge

One characteristic of modern-day cognitive psychology is a recognition that there is more than one type of knowledge and that a concern for different types of outcomes should be considered in designing appropriate learning/training environments. Although this realization is not actually a new thrust in learning research, it has taken on a rather different flavor since Bob Gagné's (e.g., 1962, 1965) seminal work 20 years ago on different types of learning and the implications of this distinction for instruction (i.e., the conditions for learning and the implications of that orientation for instruction).

Cognitive psychology acknowledges that there are different types of knowledge, most frequently distinguishing between procedural knowledge and propositional knowledge (also referred to as declarative knowledge or semantic knowledge). Procedural knowledge refers to one's knowledge of how to do the various procedures necessary to perform some task. The task can be either psychomotor or intellectual, and Gagné (1977) also identifies motor skills as a separate type of learned outcome. Propositional knowledge, on the other hand, refers to the systematic and organized body of knowledge that we have about something. There are undoubtedly other types of knowledge as well. For example, Gagné and White (1978) identify four types of organized memory structures: (a) intellectual skills, (b) networks of propositions (propositional knowledge), (c) images, and (d) episodes (autobiographical and temporal data information).

There are at least two reasons why we should be more concerned with the various types of outcomes that we are trying to achieve in training. First, the different outcomes are acquired most efficiently through the presence of different learning conditions; consequently, different types of instructional learning environments are required, depending on the type of desired outcome (Gagné, 1965, 1977). We should not help a trainee to learn a procedure in the same way that we help him or her to learn a proposition. Second, since there are qualitatively different types of outcomes, we cannot expect a person to know how to do something (procedural knowledge) just because we have taught him or her something about it (propositional knowledge), any more than we can expect a trainee to know various radio call signs after teaching him or her the names of various types of radio equipment. In order to provide the most effective training possible, we need to know both the type of outcome we are seeking and the training conditions best suited for achieving that type of outcome.

How do we go about this? There are several lines of research that might be pursued by the learning laboratory. One thing that has impressed me about some of the job descriptions and objectives currently being used in the Air Force - although it is possible that I do not fully understand the system - is that there does not seem to be any way of identifying similar types of objectives in terms of psychological factors that could be related to different types of instructional modes. The objectives seem to be primarily descriptive of the desired outcomes, rather than having been analyzed or identifed in terms of the cognitive processes or mental activities that need to be engaged or are necessary in order for an individual to perform them. Although I am not a fond lover of taxonomies, perhaps there should be some concern for the types of knowledge that we are trying to help the trainees achieve with regard to both on-the-job and end-of-training performance. We need to know more about the type of categories that are most meaningful and useful for Air Force training. Then we need to know more about the types of training that are most appropriate for acquiring the various types of outcomes. Do we, for example, teach procedures differently from the way we teach declarative knowledge? Are concepts taught differently than the paired-associate learning that one engages in when trying to achieve certain, necessary terminology? Research on both of these issues could be pursued at the learning laboratory.

A related concern that has been receiving increased attention in recent years is the attempt to describe and understand the nature of competence as reflected in differences between experts and novices in a particular field. (This concern was further expanded and related more directly to Air Force training in Glaser's paper.) The focus of this research is on qualitative differences (e.g., differences in learning strategies employed or differences in the way they approach problems) rather than quantitative differences (i.e., mere differences in the amount of knowledge possessed). Bob Glaser and others at LRDC (e.g., Chi, Glaser, & Rees, 1982; Voss, Tyler, & Yengo, 1983) have performed much of this research, although other investigators have also been active in the field (e.g., Larkin, McDermott, Simon, & Simon, 1980). This description of differences between experts and novices in a field may be a useful way for going about the task of meaningfully describing the type of knowledge or competence that we want various trainees to acquire. For example, what does a radio operator do in carrying out his or her job? What are the strategies that are used in performing this job in a competent manner? How do these strategies differ from those used by a novice and can they be specified in a manner relevant to the task of moving a person from being a novice to an expert in the way the job is performed? How we actually go about doing this is not completely clear, but one of the problems that might be addressed at the laboratory would be an attempt to apply some of the more qualitative approaches that have been used in defining the nature of competence to on-the-job performance being carried on by Air Force personnel in the field.

We also should be aware of the fact that the nature of the knowledge being acquired may change as learning progresses; as one moves through a unit or set of procedures for learning something, different types of instructional support may be needed. John Anderson (1982) has recently described the various stages one goes through in learning a procedure or procedural knowledge. He has suggested three stages that parallel Fitts' (1964) stages of learning a motor skill, although Anderson refers to the newer stages by somewhat different names.

The first of these stages Anderson refers to as the declarative stage in which the concern for declarative knowledge is paramount. It may be that in this stage we need a fair amount of instructional support in terms of mnemonics or something that would aid the learner in holding things together long enough for the core material to be integrated into a meaningful part of his or her knowledge structure. The second stage that Anderson has identified is called knowledge compilation. This stage is concerned with the transition from knowledge represented in a declarative form to knowledge represented in a procedural form and is interpretive and integrative in nature. Finally, the stage in which the skill has become autonomous and there is a gradual increase in the speed with which the skill can be performed is referred to by Anderson (1982) as the procedural stage.

If our goal is to train people to perform procedures, which seems to be a large part of the technical training conducted by the Air Force, we may need to become more concerned for the type of declarative knowledge we need to use, ensuring either its usefulness as a crutch for initial learning or its relevance to the task in a more direct manner. That is, what do we need to teach over and above the specific procedures and what is needed to support the transitions between the various stages of learning?

Individual Differences

A third concern does not strictly grow out of cognitive psychology, but it is so much at the core of current thinking about learning from instruction that it would be an oversight not to mention it. This concern is for the nature of individual differences in learning (which has taken on a decidedly cognitive flavor) and a corresponding concern for aptitude-treatment interactions (ATIs). ATIs pertain to situations in which an instructional method that is optimal for one type of student may produce poorer learning for a different type of student, with the reverse being true for a different instructional method (Cronbach & Snow, 1977). While ATI research has not been overwhelming in terms of its findings, some fairly fundamental conclusions can be stated with confidence. For example, low ability students seem to do consistently better in a training environment that has a high degree of structure, thereby placing a low informational processing load on the learner. High ability students, on the other hand, perform at a slightly lower level in that kind of an environment than they do in an instructional environment that places a high information processing load on the learner (Snow, 1977).

Put in absolute terms, it is not clear where Air Force trainees would be with regard to ability. Neither is it clear what the effects would be if one looked at prior knowledge (e.g., Tobias, 1976) rather than ability. Findings of this nature, however, may be worth pursuing. Other than that, the ATI research has not been all that fruitful. I think there are two reasons why this might be the case. First, I think that not enough consideration has been given to the various types of knowledge involved in the desired outcomes; that is, TOIs (treatment-outcome interactions) and AOIs (aptitude-outcome interactions) may be involved as well as ATIs.

The other reason has to do with the fact that to a very large extent, ATI research has used off-the-shelf tests to measure aptitudes; tests that were designed for other purposes - namely, unidimensional prediction in a more-or-less fixed instructional environment. Perhaps what we need to do, if we are serious about trying to develop a viable ATI technology, is to pursue the possibility that we need to develop completely new aptitude tests designed for differential prediction. The notion of new kinds of aptitudes obviously is not a totally fresh idea (Glaser, 1972). What I have in mind here is to go through the process of developing tests whose purpose would be one of differential prediction for different learning environments and/or instructional procedures. That is, you take different types of outcomes or different types of instructional treatments, develop a pool of test items, and then select items for the test that will discriminate appropriately. Thus, we would actually build new psychometric instruments for the purpose of differential prediction rather than unilateral prediction, which is the purpose of traditional psychometric tests.

There are several ways one might go about this task. Ideally, one needs a theory of what is and is not important in these different instructional environments and the types of mental activities that students engage in when learning a specific type of outcome in a specific type of learning environment. It would help to have a viable theory of cognition to guide this development, and I think we may have the beginnings of what might prove to be appropriate conceptions of (a) a useful cognitive model of learning, (b) the task demands of various instructional environments, and (c) the task demands of various outcomes. For example, considerable progress has been made in recent years in developing a process theory of aptitudes (e.g., Snow, 1980). This work is very encouraging in terms of developing a better understanding of aptitudes and how they relate to instruction. In general, however, these aptitudes tend to be of a general nature in the sense of cutting across various subject-matter domains and environmental (instructional) task demands.

There have been some notable and important attempts in the past 5 years to understand the nature of psychometric performance. The concern of this research has been for identifying the cognitive processes responsible for the performance observed on psychometric tests - that is, as Carroll (1976) has so aptly put it, concern for "psychometric test as cognitive tasks." Two basically different, but not mutually exclusive, paradigms have been used in this research. Pellegrino and Glaser (1979) have labeled these two paradigms the "cognitive components approach" and the "cognitive correlates approach." In the cognitive components approach, performance on parameters of complex information-processing tasks taken directly from items on psychometric tests (e.g., analogy items) are correlated with scores from the psychometric tests, on the assumption that individual differences on the component tasks underlie individual differences in overall test performance (e.g., Sternberg, 1979). In the cognitive correlates approach, parameters from the simple kinds of information-processing tasks studied in the laboratory are correlated with scores from tests of mental ability, on the assumption that individual differences in the cognitive processes represented by these tasks underlie observed scores on psychometric tests of mental abilities (e.g., Hunt, Frost, & Lunneborg, 1973). These approaches could also be used to investigate various types of complex learning tasks, such as those encountered in many types of training situations.

In another sense, because we are concerned primarily with performance on learning tasks, psychometric instruments to a very large extent are not merely attempts to understand the individual differences in psychometric ability but rather, the individual differences in learning tasks themselves. Some of my own research (e.g., Shuell, in press) and some of the research of John Bransford (e.g., Bransford, Stein, Shelton, & Owings, 1982; Bransford, et al., in press) has followed this approach. We start by taking some type of learning task (either laboratory-type tasks, which at this early stage of research is not totally unreasonable, or more complex types of tasks similar to those that might be encountered in a classroom) as the criterion task we are interested in understanding. Next we ask, "Since individuals differ in their ability to learn these tasks, is it possible to understand the nature of this differential performance through either some sort of a componential or a correlational approach?"

In some of my research (e.g., Shuell, in press), for example, we have looked at factors that might reduce some of the differences in performance between the good learners and the poor, slow learners on a free-recall task. In one study, we used categorized word lists and found that individuals identified as either fast learners or slow learners on a separate free-recall test of unrelated words profited differentially from instructions to organize the material during learning. These instructions improved the performance of the poor learners tremendously but actually had a small negative effect on the performance of the good learners. These data can be used in two ways. First, theoretically, in that perhaps one of the things differentiating good and poor learners is something analogous to a production deficiency. Slow learners may have appropriate strategies but do not use them unless they are prompted. Second, the data also can be looked at from a more practical orientation; i.e., this may be a useful way to provide instructional support. If one were to pursue this approach in the present context, one might take jobs performed in the field, look at the learning or acquisition of those jobs, and then try to identify those factors that explain individual differences in the rate of learning those jobs and how some of those differences might be reduced.

Cognitive Research on Learning

Until recently, cognitive psychology has shown very little concern for learning as such - that is, concern for the factors or variables that influence changes in human performance, knowledge structures and/or conceptions (Anderson, 1982; Greeno, 1980; Langley & Simon, 1981). Rather, cognitive psychology has focused on identifying and describing the various stages and processes that are involved when humans process information. The most significant impact that cognitive psychology has had on the psychology of learning is to change the conception of learning from a process that is passive and reactive to a process that is active and goal oriented. Although this is an extremely major and important impact in itself, there have been few attempts so far to follow up this change in conception with investigations of the variables that might be responsible for these psychological changes we call learning.

There have been, however, several attempts to articulate the role of learning from a cognitive-psychology and human-information-processing perspective. Most of the recent work has occurred within the area of artificial intelligence, where the goal has been to develop computer programs that can learn. The only formal definition of learning from this perspective that I have come across is the following one suggested by Langley and Simon (1981):

> Learning is any process that modifies a system so as to improve, more or less irreversibly, its subsequent performance of the same task or of tasks drawn from the same population. (p. 367)

In many ways, the differences between this definition and more traditional, behavioristic definitions of learning appear to involve flavor more than substance, although some very important differences do exist between the two orientations.

Probably, the first modern attempt to develop a cognitive conception of learning was Rumelhart and Norman's (1978) logical analysis of the nature of learning within a schema-based representational system. Thinking about knowledge representation and memory in terms of schemata is very common in cognitive psychology. According to Rumelhart and Norman, there are three qualitatively different kinds of learning.

1. Accretion, or the encoding of new information in terms of existing schemata.

2. Tuning or schema evolution, which involves the slow modification and refinement of a schema as a result of using it in different situations.

3. Restructuring or schema creation, which is the process whereby new schemata are created.

Although most models of memory involve learning by means of accretion, Rumelhart and Norman (1981) go on in a more recent chapter to suggest that learning which involves the creation of new schemata occurs as the result of analogical process; that is, we learn new schemata by relating new information to old schemata in analogical ways.

John Anderson (1982) has suggested that we learn procedure knowledge by making inferences from declarative (propositional) knowledge. He has developed a computer program, called ACT, capable of learning how to do geometry proofs. Briefly, the program involves an adaptive production system that engages in the types of learning referred to above as tuning. Three learning mechanisms are used as the basis of tuning:

1. Generalization, a process by which production rules become broader in their range of applicability.

2. Discrimination, a process by which production rules become narrower in their range of applicability.

3. Strengthening, a process by which better rules are strengthened and poorer rules are weakened.

These approaches represent the beginning of a concern to develop a viable cognitive theory of learning. Although they do represent a major advance, they still seem to me to be rather removed from instructional implications. An attempt to bring together the concerns of traditional learning psychology with the newer concerns of cognitive psychology is something that should be at the very focal point of the new laboratory. Considerable thought needs to be given to developing the most productive relationship possible between learning and cognition, on the one hand, and instructional procedures appropriate for Air Force training, on the other.

Research on Learning Variables That Have Instructional Relevance

The final concern that will be discussed here is for research on those learning variables that we know have important instructional implications or that have effects as far as teaching is concerned. We know a number of things that are worth pursuing further.

If we accept memory and transfer as the bottom line of any training or educational program (i.e., the true goal of training is not to have the trainees demonstrate learning in the classroom, but to have them retain and be able to use the knowledge they have acquired in novel situations), then we need to consider some of the factors that influence these goals. (See Royer's paper for a further discussion of transfer.) We know quite a lot about how to improve both memory and transfer by providing for the appropriate conditions during learning. We need to extend this knowledge, especially in the prescriptive sense discussed in the preceding section, and incorporate it into our training programs.

For example, we know that distribution of practice during learning has extremely large effects in facilitating memory (Bloom & Shuell, 1981; Shuell, 1981a), at least with verbal materials and 24-hour distribution intervals. Will these same procedures work for more meaningful materials? We have tried in our laboratory to extend these findings to both concept learning and prose learning, but so far, our results have been inconclusive for a variety of reasons that are irrelevant here.

The use of mnemonics - both visual and verbal - is another topic for which there are encouraging results. Are there situations in Air Force training where mnemonics could be used appropriately, either as a crutch on the way toward making performance automatic, or in ways that are more directly related to the task being acquired? My hunch is that mnemonics have their best effects as temporary adjuncts - that is, they serve like a glue to hold things together until the material is integrated into a larger knowledge structure.

Possibilities for Research

Research conducted in an Air Force learning research laboratory can and should take a variety of different forms and directions. Specific research projects, however, should incorporate the various types of concerns discussed earlier in this paper. For example, research focusing only on how skills are learned and/or retained is incomplete as far as the overall purpose of the laboratory is concerned; instructional variables and concern for how one might increase the effectiveness of the training program - designed to produce specific learning outcomes - should be evident in each project. To some extent, this orientation involves thinking about learning and various learning variables in a somewhat different, more cognitive manner than is the case with traditional research on learning, as well as being aware of the necessary relationship between learning and instructional variables.

This section of the paper will discuss several possible research projects and try to show how the various concerns voiced above might be integrated in a productive manner. The specific examples discussed, of course, are intended merely to be representative of the types of research that might be pursued at the laboratory.

Acquisition and Retention of Procedural Knowledge

Most learning involved in Air Force training is procedural in nature (i.e., using some intellectual or physical skill in such endeavors as operating, maintaining, or repairing a specific piece of equipment), although some declarative (propositional) knowledge (e.g., learning radio call signs, specific regulations, or provisions of the Uniform Code of Military Justice is also seen as being important in its own right. Yet, this emphasis on the performance of procedural tasks does not necessarily mean that the learning of declarative knowledge is of little importance in Air Force training.

Aside from learning declarative knowledge for its own sake, declarative knowledge plays several, potentially important roles in the learning of procedural knowledge. Anderson (1982), for example, suggests that the learning of declarative knowledge is the first stage involved in the acquisition of procedural knowledge and that we learn procedural tasks by making inferences from this declarative knowledge. This theory has a great deal of plausibility, and assuming its correctness, research is needed on the type of declarative knowledge that should be presented in support of learning the types of procedural knowledge involved in various Air Force jobs (or job components as discussed in Gagné's paper). In addition, empirical evidence is needed that is relevant to the task of determining how we can best facilitate the learning of both the declarative and the procedural knowledge, as well as the transition between them. Anderson refers to this stage as the knowledge compilation.

The presentation of declarative knowledge is not sufficient in and of itself for the desired learning to occur. Learning is an active process. What the learner does is just as important, if not more important than, what the instructor does. For example, cognitive learning theory emphasizes the crucial role played by the particular knowledge or schema that the student happens to have activated at the time learning takes place (e.g., Bransford, 1979). If an appropriate schema is activated, then the desired learning occurs much more easily than would be the case if the schema had not been activated. Consequently, various types of instructional adjuncts, aids to comprehension (e.g., Mayer, in press), and/or visual mnemonics (e.g., Levin, 1981) may prove to be useful instructional devices for increasing the efficiency of learning by helping to elicit appropriate schemata.

As far as retention is concerned, there appears to be only a very few variables that affect memory independently of any effect that the variable might have on learning (Shuell & Lee, 1976). Degree of original learning, of course, is one of these variables, and anything that increases how well something is learned will consequently have an indirect effect on how well it is remembered. Another variable is distribution of practice during learning, which generally has no effect (or even a slight decremental effect) on learning but has a powerful facilitative effect on memory (Bloom & Shuell, 1981; Shuell, 1981a). The third variable for which there is reasonably good evidence is similarity of interfering material, but the practical implications of this variable are not clearcut.

In addition, there is some suggestive, but at present rather speculative, evidence that organization (Shuell, 1981b) and the use of multiple contexts during learning (MacKenzie & White, 1982) may likewise affect memory independently of learning. The latter variable takes on added interest when one considers the evidence suggesting that transfer is increased when the learner works on a variety of tasks during learning rather than working on a single task (Duncan, 1958; Morrisett & Hovland, 1959), or when irrelevant information is present during learning (Overing & Travers, 1966, 1967).

These various considerations lead to the following research questions (among many others) concerning the acquisition and retention of procedural knowledge:

1. Do any of the following aids to comprehension facilitate the acquisition of the declarative knowledge learned during the process of acquiring procedural knowledge representative of Air Force jobs, thereby enabling the desired procedural knowledge to be learned more quickly?

   a. Advance organizers of the concrete, abstract and visual types.

   b. Visual mnemonics.

   c. Behavioral objectives.

   d. Adjunct questions of both the written and oral types.

   e. Development of "cognitive maps."

2. If instructional adjuncts such as those listed in Question No. 1 are developed for the procedural knowledge itself (rather than the declarative knowledge supporting it) and used during the "Knowledge Compilation" and/or "Procedural" stages, will the acquisition and/or retention of the desired procedure be facilitated?

3. Does the learning of procedural knowledge under conditions of distributed practice facilitate the long-term retention of the procedure?

4. Does the learning of either declarative or procedural knowledge in multiple contexts (yet self-contained in a particular training environment) facilitate the retention and/or transfer of the material or skill being acquired?

Individual Differences in Learning from Instruction

Several aspects of individual differences in learning from instruction have already been discussed in this paper and will not be repeated here. What will be presented in this section are some possible research questions based on the concerns and research discussed earlier. In a training environment such as the present one, a concern for individual differences implies at some level a concern for individualization of instruction. In some cases, however, these instructional concerns may be primarily indirect or postponed to future research and development. Although the ATI paradigm (and the extensions of it discussed in an earlier section of this paper) is the most explicit and integrated approach in this regard, research questions over and above those suggested earlier will not be developed here due to space limitations. Such development would involve a fairly extensive statement in order to do an adequate job, and it is not feasible to go into the necessary detail in this section. Consequently, the following research questions represent a somewhat selective part of the programmatic research on individual differences that should be conducted at the laboratory:

1. In what ways do individuals who are experts in performing various Air Force jobs differ from new trainees with regard to the following:

   a. The way the task demands of the job are perceived?

   b. The way the individual goes about performing the job to the best of his or her ability?

   c. Variables that differentially affect the performance of "experts," "apprentices," and "novices" (in order to determine instructional interventions that will most efficiently help new and advanced trainees to become more proficient at performing their jobs)?

2. What variables differentially affect the performance of "good" and "poor" learners at various levels of competence, such as "novices" and "apprentices" (so that training procedures can be developed that will enable all personnel to learn and retain assigned tasks to the maximum level possible)?

Conclusion

Several themes of current research on cognitive psychology and learning from instruction have been discussed within the context of Air Force training. These themes will undoubtedly influence the research conducted by the proposed learning research laboratory and should be considered in planning the research agenda for the laboratory. This paper has tried to identify some of the topics that might be incorporated into various research projects. For example, further investigation is needed on the instructional factors that help to maintain and support the retention of skills and transfer; but we need to relate these factors to the various objectives of Air Force training and try to analyze on-the-job types of experiences, not only in terms of logical categories, but also in terms of psychological categories that have relevance for learning from instruction.

REFERENCES

Anderson, J. R. (1982). Acquisition of cognitive skill. Psychological Review, 89, 369, 406.

Anderson, R. C., & Biddle, W. B. (1975). On asking people questions about what they are reading. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 9). New York: Academic Press.

Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. Journal of Educational Research, 74, 245-248.

Bransford, J. D. (1979). Human cognition: Learning, understanding and remembering. Belmont, CA: Wadsworth.

Bransford, J. D., Stein, B. S., Shelton, T. S., & Owings, R. A. (1982). Cognition and adaptation: The importance of learning to learn. In J. Harvey (Ed.), Cognition, social behavior and the environment. Hillsdale, NJ: Erlbaum.

Bransford, J. D., Stein, B. S., Vye, N. J., Franks, J. J., Auble, P. M., Mezynski, K. J., & Perfetto, G. A. (in press). Differences in approaches to learning: An overview. Journal of Experimental Psychology: General.

Brophy, J. E. (1979). Teacher behavior and its effects. Journal of Educational Psychology, 71, 733-750.

Brophy, J., & Good, T. L. (in press). Teacher behavior and student learning. In M. C. Wittrock (Ed.), Third handbook of research on teaching.

Carroll, J. B. (1976). Psychometric tests as cognitive tasks: A new "structure of intellect." In L. B. Resnick (Ed.), The nature of intelligence. Hillsdale, NJ: Erlbaum.

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R.J. Sternberg (Ed.), Advances in the psychology of human intelligence. Hillsdale, NJ: Erlbaum.

Cooper, L. A. (1982). Strategies for visual comparison and representation: Individual differences. In R.J. Sternberg (Ed.), Advances in the psychology of human intelligence. Hillsdale, NJ: Erlbaum.

Cronbach, L. J., & Snow, R. E. (1977). Aptitudes and instructional methods: A handbook for research on interactions. New York: Irvington.

Duncan, C. P. (1958). Transfer after training with single versus multiple tasks. Journal of Experimental Psychology, 55, 63-72.

Fitts, P. M. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), Categories of human learning. New York: Academic Press.

Gagné, R. M. (1962). The acquisition of knowledge. Psychological Review, 69, 355-365.

Gagné, R. M. (1965). The conditions of learning. New York: Holt, Rinehart and Winston.

Gagné, R. M. (1977). The conditions of learning (3rd ed.). New York: Holt, Rinehart and Winston.

Gagné, R. M., & White, R. T. (1978). Memory structures and learning outcomes. Review of Educational Research, 48, 187-222.

Glaser, R. (1972). Individuals and learning: The new aptitudes. Educational Researcher, 1(6), 5-13.

Glaser, R. (1973). Educational psychology and education. American Psychologist, 28, 557-566.

Greeno, J. G. (1980). Psychology of learning, 1960-1980: One participant's observations. American Psychologist, 35, 713-728.

Hilgard, E. R. (1964). A perspective on the relationship between learning theory and educational practices. In E. R. Hilgard (Ed.), Theories of learning and instruction, The sixty-third yearbook of the National Society for the Study of Education, Part I. Chicago: University of Chicago Press.

Hunt, E. B., Frost, N., & Lunneborg, C. (1973). Individual differences in cognition: A new approach to intelligence. In G. B. Bower (Ed.), The psychology of learning and motivation (Vol. 7). New York: Academic Press.

Langley, P., & Simon, H. A. (1981). The central role of learning in cognition. In J. R. Anderson (Ed.), Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.

Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. Science, 208, 1335-1342.

Levin, J. R. (1981). The mnemonic '80s: Keywords in the classroom. Educational Psychologist, 16, 65-82.

MacKenzie, A. A., & White, R. T. (1982). Fieldwork in geography and long-term memory structures. American Educational Research Journal, 19, 623-632.

Mayer, R. E. (in press). Aids to text comprehension. Educational Psychologist.

Morrisett, L., & Hovland, C. I. (1959). A comparison of three varieties of training in human problem solving. Journal of Experimental Psychology, 58, 52-55.

Overing, R. L. R., & Travers, R. M. W. (1966). Effect upon transfer of variations in training conditions. Journal of Educational Psychology, 57, 179-188.

Overing, R. L. R., & Travers, R. M. W. (1967). Variation in the amount of irrelevant cues in training and test conditions and the effect upon transfer. Journal of Educational Psychology, 58, 62-68.

Pellegrino, J. W., & Glaser, R. (1979). Cognitive correlates and components in the analysis of individual differences. In R. J. Sternberg & D. K. Detterman (Eds.), Human intelligence: Perspectives on its theory and measurement. Norwood, NJ: Ablex.

Reder, L. M. (1980). The role of elaboration in the comprehension and retention of prose: A critical review. Review of Educational Research, 50, 5-53.

Rumelhart, D. E., & Norman, D. A. (1978). Accretion, tuning, and restructuring: Three modes of learning. In J. W. Cotton & R. Klatzky (Eds.), Semantic factors in cognition. Hillsdale, NJ: Erlbaum.

Rumelhart, D. E., & Norman, D. A. (1981). Analogical processes in learning. In J. R. Anderson (Ed.), Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.

Shuell, T. J. (1981a). Distribution of practice and retroactive inhibition in free-recall learning. Psychological Record, 31, 589-598.

Shuell, T. J. (1981b, April). Toward a model of learning from instruction. In K. Block (Chair), Psychological theory and educational practices: Is it possible to bridge the gap? Symposium presented at the meeting of the American Educational Research Association, Los Angeles.

Shuell, T. J. (1982). Developing a viable link between scientific psychology and educational practices. Instructional Science, 11, 155-167.

Shuell, T. J. (in press). The effect of instructions to organize for good and poor learners. Intelligence.

Shuell, T. J., & Lee, C. Z. (1976). Learning and instruction. Monterey, CA: Brooks/Cole.

Simon, H. A. (1981). The sciences of the artificial (2nd ed.). Cambridge, MA: M.I.T. Press.

Smiley, S. S., Oakley, D. D., Worthen, D., Campione, J. G., & Brown, A. L. (1977). Recall of thematically relevant material by adolescent good and poor readers as a function of written versus oral presentation. Journal of Educational Psychology, 69, 381-387.

Snow, R. E. (1977). Research on aptitudes: A progress report. In L. S. Shulman (Ed.), Review of research in education 4, 1976. Itasca, IL: Peacock.

Snow, R. E. (1980). Aptitude processes. In R. E. Snow, P. A. Federico, & W. E. Montague (Eds.), Aptitude, learning, and instruction (Vol. 1): Cognitive process analyses of aptitude. Hillsdale, NJ: Erlbaum.

Spiro, R. J., Bruce, B. C., & Brewer, W. F. (Eds.). (1980). Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artificial intelligence, and education. Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1979). The nature of mental abilities. American Psychologist, 34, 214-230.

Tobias, S. (1976). Achievement treatment interaction. Review of Educational Research, 46, 61-74.

Voss, J. F., Tyler, S. W., & Yengo, L. A. (1983). Individual differences in the solving of social studies problems. In R. F. Dillon & R. R. Schneck (Eds.), Individual differences in cognition. New York: Academic Press.

Wittrock, M. C. (1983, April). Generative reading comprehension. Paper presented at the meeting of the American Educational Research Association, Montreal, Quebec, Canada.

CHAPTER 6

APPROACHES TO THE ACQUISITION OF UNDERSTANDING

James M. Royer
University of Massachusetts, Amherst

This paper outlines a strategy for the development of instructional procedures that will assure students have understood the instruction they have received. The paper begins with a description of a taxonomy of educational problems, followed by a discussion of what it means to understand from the perspective of cognitive theory. This is followed by the suggestion that satisfactory performance on transfer objectives will provide an index of the extent to which students have understood the instruction they have received. Given the acceptance of the transfer objective as an operational definition of understanding, the paper then goes on to describe four techniques that facilitate the understanding of instruction. This section is followed by a description of research relevant to keeping students involved in learning tasks and by a section describing the importance of assessing knowledge in specific content areas. The next section of the paper presents a revision of the ISD model that incorporates many of the suggestions contained in this paper. The final section of the paper contains suggestions for specific research projects that the Air Force could undertake in a learning research laboratory.

A Taxonomy of Educational Problems

By nature, I am the sort of person who attempts to impose some boundaries and structure on a problem before I can think it through. Hence, I began my task on this project by setting up a kind of classification system that is similar to one I and others have used previously (e.g., Rothkopf, 1968; Royer & Allan, 1978). The gist of this system is that educational problems can be divided into three types: problems involving observable student behaviors, problems involving the mastery of basic, low-level information, and problems involving the understanding of more complex materials.

Problems involving observable behaviors most often entail getting students to engage in task-appropriate behavior and getting them not to engage in task-inappropriate behavior. Most problems of this sort can, I believe, be successfully dealt with through the application of behavior modification principles and/or through the use of task analysis procedures.

The second type of problem entails the mastery of basic low-level performance. I have in mind here the mastery of low-level terms, definitions, quantitative entities, and concepts that provide the foundation of any area of study. Here again, I think we have powerful psychological machinery that can be activated to assist in acquiring information of this sort. I think associative learning theory can tell us a great deal about setting up ideal conditions for the mastery of basic information. Associative learning theory is out of vogue now, but we should not forget that it tells us a good deal about practice schedules, review intervals, and conditions of transfer, all of which are valuable components of an instructional sequence designed to assure mastery of basic information. If we embed the principles of associative learning theory within a design procedure for the systematic production of instructional materials, such as the familiar one in Figure 6-1, we have a particularly powerful method of developing effective instruction. The principles of associative learning theory provide the guidelines for instructional development, and the systems model provides the framework under which this development occurs.

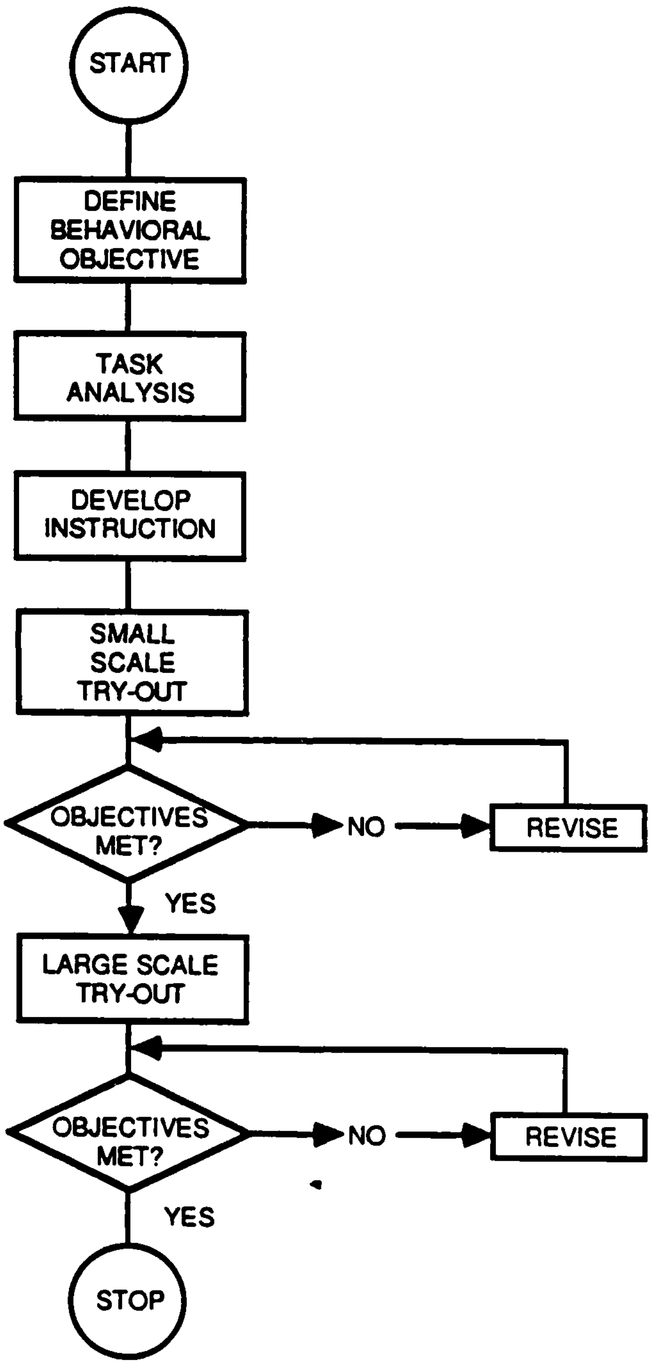


Figure 6-1. Traditional ISD Model.

We now turn to the third of my categories - problems involving understanding. Understanding is one of those terms that the early proponents of behavioral objectives urged us to purge from our vocabularies because it was so hopelessly vague that it could never serve as a meaningful educational goal (Mager, 1962); but our understanding of what it means to understand has increased considerably in recent years, principally because the question of what it means to understand is at the very heart of rapidly developing cognitive theory.

Understanding from a Cognitive Perspective

My favorite way of illustrating the change in the way we think about the process of understanding is to invoke the contrasting metaphors the linguist Michael Reddy has used to illustrate the erroneous implicit belief we have about meaning in language. Reddy (1979) has pointed out that the way we describe meaning documents the existence of an unconscious acceptance of what he has called the conduit metaphor of meaning in language. Consider the following examples which are merely a smattering of the hundreds of examples Reddy has collected. "Did you catch what I meant?" "Your ideas just don't come across in your writing." "Your speech effectively conveyed your beliefs." "Did you get the author's message?"

Reddy has argued that the examples above reflect the acceptance of a model of meaning that says that the sender of a message packages the meaning of the message in the form of speech or writing, sends that message through some communication conduit, whereupon the receiver of the message unpackages it, and extracts the meaning. Reddy argues that acceptance of the conduit metaphor means one accepts the position that the meaning of a message resides within the message itself. Incidently, the belief that meaning resides within a message is an integral part of associative learning theory. Within the framework of associative learning theory, memory is merely the blackboard that experience writes on.

Reddy argued that the conduit metaphor was wrong, and he offered a substitute that he called "the tool-maker metaphor," which he believed to more accurately represent the process of deriving meaning. Reddy began his metaphor with the suggestion that communication between individuals was analogous to interactions between people in slightly different environments. This situation could be imagined by thinking of people who live in separate compartments of a huge compound that is shaped like a wagon wheel. He suggested that we imagine each wedge-shaped sector of the wheel is an environment that has much in common with one another - water, trees, plants, rocks, and the like - but none of the environments is exactly alike. The trees are different; one environment is swampy and another is dry and rocky, and so on. At the hub of the wheel there is a device that delivers sheets of paper to people in the other environments. This is the only means of communication; the people cannot communicate directly, and they cannot experience each other's environment. Reddy then suggests that one person, say the person living in environment A, discovers how to build a wooden rake to clear dead leaves and other debris away from his cultivated plants and that he decides to share his discovery with the residents of the other environments. He draws the plans for his rake on a sheet of paper and drops the plans through the slots for delivery to the people in environments B, C, and D. The people living in the other environments receive the sheets and begin to construct what they can from them.

Environment A contains a great many trees, which is why the rake was made of wood and was useful for gathering fallen leaves. The environment in sector B, however, is mostly rock and contains few trees. The person living in environment B uses rock in all of his construction, and after not being able to find a piece of wood suitable for a handle, he begins to construct a rake head out of stone. The resulting implement is too heavy to be lifted; moreover, he cannot imagine any possible use for it. Suddenly, he gets an idea. If the stone head is modified so that it consists of only two large prongs, it will be useful in digging stones from the field. Person B makes his new implement and is very happy with the result.

Environment C is wet and swampy, and the person living there cannot see any use for the device invented by person A. However, after some experimenting, he devises a hoe which is very useful for cutting the rampant plant growth that threatens his garden.

Reddy continues on with his story, but at this point we can see the direction he is taking. The individuals exist in environments that are similar enough that they recognize the general purpose of the rake plans, and each individual has received exactly the same message. However, the interpretation of the message differs because of the differing environments. A rake becomes a stone-picker or a hoe because these are the only tools that make sense in terms of environments B and C.

Reddy suggests that the nature of the language communication process is more similar to the tool-maker metaphor than it is to the conduit metaphor. We do not "receive" the meaning of a message passed along some conduit. Instead, we must construct a meaning by interpreting a message in light of our knowledge (cf. Bransford & McCarrell, 1974). Thus, the process of understanding involves an interaction between a received linguistic message, the context in which a message is received, and the knowledge background of the person receiving the message (Jenkins, 1979).

The constructivist interpretation is at the core of the cognitive explanations of understanding. Something is understood when it has been integrated in a meaningful way into the learner's existing knowledge structure. When the learner does not have any relevant knowledge that he can use to construct an interpretation, memorization may occur, but understanding will not.

The argument that understanding involves the integration of new material into an existing knowledge structure is an interesting theoretical statement; and, as I will soon argue, it provides guidelines for developing approaches to enhance understanding, but it leaves open the question of how we will know when a student has understood what we are trying to teach. I would like to suggest a slightly different type of instructional objective as the answer to this question.

Transfer Objectives as an Index of Understanding

One powerful element of the technology of instruction approach popular during the 1960s was the use of behavioral objectives as a standard for instructional success. An instructional approach was deemed successful when students receiving the instruction performed in the manner specified in the instructional objective. I think we should adopt a similar standard of performance to evaluate the success of instructional programs based on cognitive research and theory. The nature of the objectives will, however, be different. A common standard for evaluating whether a student has understood something is whether the student can use what has been learned. By use, I mean the demonstration that the student can perform some meaningful activity beyond the replication of the instruction as experienced. Thus, repeating or recognizing an aspect of instruction would not constitute evidence of understanding, whereas using the instruction as a tool for the acquisition of further instruction, or successfully using the instruction in an unfamiliar context, would be evidence that the instruction has been understood.

Thus, I am suggesting that objectives which index the transfer of instruction serve as the standard for judging the success of instructional approaches designed to enhance understanding. Transfer objectives can range from specifications that differ only slightly from the conditions of original instruction to those that have only conceptual relations to the instructions the students received. An example of the former kind of transfer task is Anderson's (1972) recommendation that comprehension of instruction be assessed by determining if students can detect paraphrases of content experienced during instruction. An example of a more distant transfer objective would be to design a specific electronic component after completing a course in basic electronics.

Approaches to Enhancing Understanding

Given that the goal of instructional efforts based on cognitive theory is to develop understanding, and given that understanding can be assessed by determining if instruction transfers, the next question is, "What does cognitive research and theory provide in the way of guidelines for developing instruction that will encourage understanding?" First, there is a very general answer to this question: Understanding will be enhanced when the instructional material can be integrated into the learner's existing knowledge structure. However, this general answer breaks down into three more specific answers. The first of these is that understanding can be enhanced by increasing the likelihood that students will bring appropriate prior knowledge to bear when receiving instruction. Second, understanding can be enhanced by providing the student with new information that will assist in integrating the instructional material into that student's existing knowledge structure. Third, understanding can be enhanced by teaching students beneficial learning strategies and by assuring that the students use the strategies. I will now discuss each of these specific answers in turn.

Activating Existing Prior Knowledge

The first approach to encouraging understanding is to activate knowledge that the students already have, so that incoming information can then be integrated into the activated structure. One of the simplest techniques for doing this is the use of mnemonics. The most thoroughly documented educational use of mnemonic procedures is a technique called the keyword method.

The keyword method as an example of a mnemonic technique. The term "keyword method" was coined by Richard Atkinson (Atkinson, 1975) to describe the technique of foreign language vocabulary learning that involves associating the foreign word with a well-known, concrete English word, which is called the keyword. The keyword need not have a meaning relationship with the foreign word. It must, however, be similar in sound. For instance, the keyword used by Raugh and Atkinson (1975) to teach the Spanish word for duck (pato - pronounced pot-o) was "pot." As another example, the keyword for learning the Spanish word for horse (caballo - pronounced cob-eye-yo) was "eye."

The keyword method divides vocabulary learning into two stages. The first step requires the student to associate the spoken foreign word with the keyword. This association is generally learned quickly because of the sound similarity between the foreign word and the keyword. The second stage requires the subject to form a mental image of the keyword "interacting" with the English translation of the foreign word. For example, the student might form an image of a duck with a pot on its head as an aid in learning the meaning of "pato" or, another example, to imagine a horse with a large eye in the middle of its forehead as an aid in learning the meaning of "caballo."

The keyword technique has proven to be a remarkably effective technique for acquiring foreign language vocabulary words in a variety of settings and with a number of different kinds of students. Atkinson and Raugh (1975) had one group of Stanford University undergraduates study a 120-word Russian vocabulary list using the keyword method while a control group, instructed to learn the words in any way they chose, studied the same list. At the end of a set period of time, the group using the keyword technique had learned an average of 55 of the words. In addition, this learning advantage persisted over time. When tested 6 weeks later, the keyword group remembered the meaning of 43% of the 120 words, whereas the control group remembered only 28% of the words. In another study involving college students, Raugh and Atkinson (1975) reported that a group using the keyword method mastered 88% of a 60-word Spanish vocabulary list, whereas a control group studying the list for the same period of time mastered only 28% of the list.

The keyword method has also proven to be a very effective means of learning foreign language vocabulary with much younger students. Pressley and Levin (1978), and Levin, Pressley, McCormick, Miller, and Shriberg (1979), have reported a number of experiments that involve teaching children in elementary school the meaning of Spanish words. In every study involving children in this age group, those students who learned the words using the keyword method learned the meaning of more words than did students who used their own methods to learn the words. Moreover, the margin of advantage for the keyword students was frequently greater than 2:1.

Thus far, the reviewed research has focused on the acquisition of foreign words. We will now consider evidence that not only does the keyword method produce superior acquisition, it also meets the criterion for understood learning in that it produces superior transfer. Pressley, Levin, and Miller (1980) reported several studies in which college students learned the definitions of infrequently used English words using either the keyword technique or a technique of the student's own devising. The students then performed a variety of tasks that included identifying when a targeted word was used correctly or incorrectly in a sentence, selecting the correct sentence into which a targeted word could be inserted, and constructing a sentence that involved the correct usage of a targeted word. In each of these transfer tasks, the group that learned the words using the keyword technique performed at a level superior to that of the control group.

Using mnemonics represents one of the simplest means of activating prior knowledge as an aid in achieving meaningful learning of new material. The research reviewed focuses on the acquisition of vocabulary words, but it should be noted that the technique could be used for acquiring a variety of kinds of information such as technical names for objects or other things, and the names of people or places (Lorayne & Lucas, 1974).

<u>Using analogies to facilitate subsequent learning</u>. Another way of activating prior knowledge that can then be used to facilitate subsequent learning is through the use of analogies. Royer and his associates (Royer & Cable, 1975, 1976) have reported several studies using this technique. They wrote two passages, one of which was concerned with the transfer of electricity through metals; the other was concerned with the transfer of heat through metals. Each of these passages was then rewritten in two forms. One form - called the "concrete" form - was written in an easily understandable form that made extensive use of analogies to assist in understanding difficult concepts. The other version - called the "abstract" version - was purposely written in a highly abstract style that was supposed to be difficult to understand.

Royer and Cable (1975) chose heat and electrical conduction through metals as topics for their passages because of the considerable similarities between the two phenomena. Metals are excellent conductors of both heat and electricity because they are crystalline in structure, and they have a great many free-floating electrons. These properties were explained by analogy in the following way:

> Heat transfer actually involves the transfer of molecular motion. In the case of heat conduction through metals, this transfer of motion occurs through a solid substance such as a bar of iron. If we were able to examine a bar of iron through an extremely powerful microscope, we would see that the interior consists of a series of regularly shaped and spaced structural units known as crystal lattices. In order to picture these lattices, imagine a box made of many Tinker Toys with smaller boxes inside consisting of other joined Tinker Toys. The solid round parts of the Tinker Toys would correspond to the molecules within the crystal lattice. In our bar of iron, which is a good conductor of heat, each of the bonded molecules within the lattice has associated with it several "free-floating" electrons. Each crystal lattice then is an orderly array of molecules surrounded by a cloud of electrons that are not attached to any particular molecule, but are free to move at random through the lattice. You can picture this by imagining many tiny particles floating through the series of Tinker Toy boxes.

By using analogies such as the Tinker Toy analogy in the passage above, Royer and Cable (1975) hoped to activate prior knowledge that could be used to facilitate the understanding of the material being presented in the passages. This understanding could then be assessed by presenting the students with related material and seeing if what they had learned previously eased the learning of the related material; in short, determining whether the learned material would positively transfer to the learning of additional material.

The results of an experiment supported the prediction that activating knowledge by using an analogy would ease the learning of subsequent related material. Students who read the analogy material prior to reading a second passage learned at least 40% more from the second passage than did a group that did not read the analogy material.

The results of the initial Royer and Cable (1975) study have been replicated many times (Perkins, 1978; Royer & Cable, 1976; Royer & Perkins, 1977). In addition, Royer and Perkins (1977) were able to demonstrate that the meaningful learning produced by using analogies persisted over time. They presented college students with an initial passage with analogies, or with a control passage; and they then had the students read a second transfer passage immediately, after 2 days, or after 1 week. Their results indicated that the positive transfer attributable to using analogies was as strong after 1 week as it was immediately.

Coming up with exactly the right analogy to achieve meaningful learning of conceptually difficult material is a truly creative act. This is especially true since many analogies do not work very well. Describing a living organism, for example, as being like a factory that takes in raw materials and puts out waste products may capture one attribute of living organisms; but the "fit" of the analogy is so poor that it may generate so much confusion that it negates any benefits that might occur with its use. For example, a discerning student might point out that one of the defining features of factories is that they produce a product. If living organisms are like factories, what "products" do they produce?

Another caution with using analogies to assist in achieving meaningful learning is that the supposedly known part of the analogy must, in fact, be part of the learner's knowledge. If we told you that Aplysia (a large sea snail) is a favorite organism of study by neurobiologists because it has a characteristic of its nervous system that is similar to the also favored squid, chances are that we would not have enlightened you in the least. Only if you know that squid have very large nerve cells, thereby making them easy to examine and study, would you benefit from being told that Aplysia is a favorite organism of study for the same reason. Teachers should be very careful to make sure that the supposedly known part of the analogy is, in fact, part of the learner's knowledge. Otherwise, using the analogy certainly will not help the learner, and, in fact, may be worse than nothing in that learners now have two things they do not understand rather than one.

Questions such as the ones mentioned above are very fruitful issues for future researchers. Moreover, analogy research is receiving increasing attention in both the basic and applied literatures. An example of the former is the insightful analysis recently presented by Gick and Holyoak (1983). On the more applied side, a paper written by a group of Dutch psychologists examines the educational uses of analogies and will soon appear in the Journal of Educational Psychology.

Enhancing Meaningful Learning by Teaching Students Information that Will Ease Subsequent Learning

In the previous section, I discussed techniques for achieving meaningful learning by activating information the student has already learned. In this section, I examine techniques for achieving meaningful learning that involves teaching the student new information that can then be used to ease subsequent learning.

Advance organizers. In the early 1960s, David Ausubel (1960, 1963) suggested advance organizers as a device for easing the learning of subsequently presented material. Advance organizers are statements that are "presented at a higher level of abstraction, generality, and inclusiveness than the new material to be learned" (Ausubel, Novak, & Hanesion, 1978, p. 171). Ausubel envisioned advance organizers as a bridge between "what the learner already knows and what he needs to know before he can meaningfully learn the task at hand" (Ausubel, Novak, & Hanesion, 1978, p. 172).

Ausubel and his fellow researchers conducted many studies during the 1960s (Ausubel, 1960; Ausubel & Fitzgerald, 1961, 1962; Ausubel & Youssef, 1963; Fitzgerald & Ausubel, 1963) that demonstrated the efficacy of presenting students with advance organizers prior to engaging in a learning task. A typical study by Ausubel and Youssef (1963) involved having college students read a passage on Buddhism after reading a "comparative advance organizer" (examining the relation between Buddhism and Christianity) or a control passage. They found that the students who had initially read the organizer passage retained more from the Buddhism passage than did the students who initially read the control passage. In addition to Ausubel's studies which have shown greater acquisition of subsequently presented material, other studies have shown beneficial effects of advance organizers using more remote transfer tasks. Grotelueschen and Sjogren (1968), Merrill and Stolurow (1966), and Scandura and Wells (1967) have all shown that material preceded by advance organizers is more likely to transfer to remote tasks than material not preceded by advance organizers.

At this point, there seems to be little doubt that advance organizers can improve the learning of subsequent material under certain conditions. Ausubel (Ausubel, Novak, & Hanesion, 1978) has indicated two of these conditions. First, the material to be presented must involve unitary topics or related sets of ideas. If the ideas to be presented are highly varied, as might be the case, for example, in a college level survey class (e.g., introductory psychology), then it would not be possible to generate a single advance organizer to encompass all of the ideas. This does not preclude, however, breaking the subject matter into unitary topics and then preparing advance organizers for each topic.

Ausubel's second circumstance is that a true advance organizer must be developed given knowledge of the learner's existing knowledge structure. This means that if advance organizers are to serve as a bridge between material to be presented and material the learner already knows, it is essential, in fact, to know what the learner already knows. Otherwise, a teacher might end up presenting a learner with two things (the advance organizer and the target material) that the student has difficulty in learning rather than just one (the target material).

Models as a form of concrete advance organizer. Ausubel defined an advance organizer as a statement that was at a higher level of abstraction than the target material to follow. Another kind of advance organizer that has received considerable investigation is the "model" or "concrete" advance organizer.

A model or concrete advance organizer is either a statement or an illustration that provides a concrete general overview of a system the student is going to study in detail. An example of the use of a model as an advance organizer can be found in Mayer's research on teaching college students a computer programming language (Mayer, 1975, 1976; Mayer & Bromage, 1980). Mayer (1975) presented his "model" students with a text that began with a diagram of the inside working components of a computer. The diagram included an "input window" (described as a ticket window) and "output pad" (described as a pad of message paper), a "memory scoreboard" (described as an eight-space, erasable scoreboard), and a "program list with pointer arrow" (described as a shopping list). The use of descriptions such as a "ticket window" for the input window was designed to make the unfamiliar components more familiar by relating them to information the students already possessed. After the students had become familiar with the model, they received instruction on computer programming that made continued reference to the model. In effect, the student was asked to "role play" what the computer did with each of the programming statements.

Another group of students in Mayer's (1975) study were instructed on computer programming in a more traditional manner. They were given definitions of programming statements and examples of the appropriate use of each of the statements. Following instruction, each of the groups was asked to write a number of short computer programs that varied in type of problem and in complexity. A comparison of the performance of the groups revealed that students who had received the model as an advance organizer performed significantly better on the programming transfer task than did students who received the traditional instruction. The results were subsequently replicated in later studies (Mayer, 1976; Mayer & Bromage, 1980).

Mayer's research on computer programming represents one instance where a system or process model can serve as a highly beneficial advance organizer. In fact, models can be beneficial in any area of study involving a concrete system. Imagine, for example, that you were teaching an automotive repair course. A good strategy would be to begin the course with a model overview of the functioning of an automobile. Moreover, it would probably be beneficial to review the model as the students move from one system to another. So, for example, if the students had been working on the carburetor and were ready to move to the ignition system, it would probably be worthwhile to present the model of the entire car again so that they could relate the functioning of the ignition system to the operation of the entire car.

Techniques for Maintaining Active Cognitive Processing

In the sections above, I discussed what it meant to understand, and techniques for enhancing understanding. However, there is another problem in the general area of instructing for understanding. This is the problem of how to maintain an appropriate level of cognitive processing and how to focus attention on the instructional content that is believed to be most important (cf. Royer, Bates, & Konold, 1983).

Giving students instructional objectives. When instructors teach a course, they do not expect the student to master all of the information presented. Instead, they generally have specific goals that they would like the students to attain, and only part of the information presented will be directly relevant to those goals. The material not directly relevant to the goals is presented for its interest value, as examples of the principles or rules described in the goals, or as context materials that will make the targeted material easier to understand.

However, students sometimes have difficulty in distinguishing between information that is important for them to master and material that is peripheral to the central objectives of instruction. They find themselves mastering examples and context material, while failing to attend to the most important information.

The most direct way of dealing with the problem of student uncertainty about what to focus on is to provide students with the instructional objectives for the course. In general, these objectives should be either the same as, or derived from, the objectives that the teacher or instructional developer began with prior to designing the instructional sequence.

In the last 10 years, there has been a considerable amount of research on the effects of providing students with instructional objectives. This research began with a study by Rothkopf and Kaplan (1972) in which college and high school students either read or did not read learning objectives prior to studying a 3,000-word text concerned with systems training at a telephone company. They found, as one might suspect, that students who received objectives learned more of the material specified in those objectives than did students who did not receive objectives. Much more surprising, however, was the fact that the group that received objectives also learned more of the material that was not keyed to an objective. This result suggested that providing students with learning objectives has a general beneficial effect in addition to the specific effect of promoting better learning of material identified in the objectives. The results of the Rothkopf and Kaplan (1972) study were subsequently replicated and expanded in studies by Kaplan (1976), Kaplan and Rothkopf (1974), Kaplan and Simmons (1974), and Nassif-Royer (1977). In addition, Duchastel and Merrill (1973) have provided a review of the early research in this area.

There is little doubt that providing students with learning objectives can have positive effects. There is a danger, however, that teachers should be aware of. Ernst Rothkopf (1968) once suggested that many students operate in accordance with the "law of least effort." That is, they expend only the amount of learning effort that is required to attain instructional goals. Teachers also can be guilty of operating in accordance with the law of least effort, and this is particularly bad when they generate and distribute hastily prepared learning objectives.

It is very easy to generate trivial learning objectives. You simply say that at the end of instruction, students will know such and such facts. The danger with trivial objectives is that they focus attention on trivial, low-level knowledge that has little utility outside the classroom. However, if teachers expend the time and effort required to generate good cognitive objectives, the students will benefit from knowing what aspects of instruction to focus on, and the instruction process, as a whole, will benefit from a focus on goals that are truly important.

Asking questions during instruction. Another technique for maintaining active cognitive processing is to ask students questions during instruction. We can see when students are looking at a book or watching a teacher, but we cannot see whether they are doing anything besides merely gazing at the book or teacher. That is, are they, in fact, processing the information they are receiving? Asking questions is one way of increasing the likelihood that a student is actively attending to instruction.

Inserting questions in written material. Several years ago, Anderson and Biddle (1975) conducted an extensive review of the research literature on questioning while people read; they found research from early in this century (Gates, 1917) that showed increased learning and retention as a function of questioning. In the last 15 years, there has been a veritable explosion in research on this topic, and Anderson and Biddle's (1975) review has been brought up to date with two recent reviews by Rickards (1979) and Andre (1979). The bulk of this research can be traced back to a study by Rothkopf (1966), and we will examine this experiment in detail as an exemplar of much of the research that followed.

Rothkopf (1966) had students read a 5,200-word text concerned with oceanography, under several different questioning conditions. Thirty-nine objective, completion-type questions were generated from this text; and some of the students received these questions prior to reading the text. Notice that this procedure would allow the students to search for content relevant to answering the questions they had just seen. Other students received a set of five or six questions immediately after reading the segment of text in which the answers to the questions could be found. In other words, this group knew they were going to receive questions, but they did not know what the questions were until they had finished reading the relevant segment of text. Two final groups were control groups that read the text after having been instructed to learn as much of the content as they could.

After all of the students had read the text, a test was administered that contained two kinds of information. The first kind was information asked for in the questions given either before or after reading the text. The second kind of information was information that had not been asked for in any previous questions. As one might expect, students who had received questions either before or after reading the text performed better on the same questions when they appeared on the test than did students who had not received the questions. Thus, questioning students served to increase the learning of that material asked for in the questions. The second major outcome of Rothkopf's study was not nearly as predictable. The students who received questions after reading the relevant text materials recalled more of the material not tested in the questions than did any of the other groups. The result suggests that students who knew they were going to encounter questions after reading a segment of text attended to all of the text more carefully than students in most of the other groups, thereby improving their performance on materials relevant to answering questions and material that would not be tested until the final post-test. In other words, inserting questions in text increased the extent to which that text was actively processed by the students.

The beneficial effects associated with embedding questions in text have been replicated many times since Rothkopf's (1966) study. For instance, the three review articles previously mentioned (Andre, 1979; Anderson & Biddle, 1975; Rickards, 1979) examined over 100 separate studies, the vast majority of which reported results consistent with Rothkopf's (1966) original study.

Choosing the right level of inserted questions. Another important concern about embedding questions in text regards the level at which the questions should be written. Should questions call for specific factual information, or should they be written at a higher level, calling for the application of learned information to instances not encountered in the text? Watts and Anderson (1971) reported a study that addressed this issue. They had high school students read a text, about psychologists and their theories, under four question-type conditions. Two groups read the passages and then answered questions that asked for a low-level factual recall or the selection of an example of a principle that had been seen before. A third group was asked application questions that called for applying principles to examples they had not seen before. The final group read the passages without questions. After completing the study phase of the experiment, all of the students took a post-test that contained factual and application questions, some of which had been seen before the question groups and some of which were new. The results of the study were that the students receiving application questions performed better on the test than did any of the other groups. The application group not only performed at a higher level on application questions, but they also performed at a very high level on the remaining types of questions.

A learning strategy approach based on cognitive theory. Up to this point, we have focused on isolated techniques for focusing student attention and maintaining high levels of cognitive processing. Donald Dansereau and his colleagues recently combined a variety of techniques into a cognitive-based learning system (Dansereau, 1978; Dansereau, Actkinson, Long, & McDonald, 1974; Dansereau, et al., 1979). Dansereau's system is designed to teach students strategies that will enhance the learning and comprehension of instruction to which they are exposed.

The steps in Dansereau's system are encoded in the acronym "MURDER." The M stands for establishing an appropriate mood for study; the U, for promoting understanding which involves activities such as marking important and different ideas; the R, for recalling material without referring to the text; the D, for digesting material by amplifying it and storing it; the E, for expanding knowledge by self-inquiry; and finally, the R, for reviewing mistakes by learning from tests.

Each of the steps in Dansereau's system involves a number of substrategies. For example, substrategies taught to students under the recall step include paraphrasing of incoming text (to transform it to the student's own words) and developing imagery of the material whenever it is possible to do so; networking of text material that involves linking key concepts together in a manner that explicates hierarchical and semantic relationships between the concepts; and finally, analysis of key ideas, in which the student identifies key ideas or concepts, develops systematic definitions and elaborations of those concepts, and then interrelates the important concepts. Dansereau and his colleagues have developed a two-semester-hour course at Texas Christian University that teaches students to utilize the strategies in his study system. Students who have completed this course have performed significantly better on end-of-semester learning exercises than did students who did not take the course.

Accessing Prior Knowledge

The constructivist view of cognitive learning places strong emphasis on the role of prior knowledge. If the learner does not have prior knowledge relevant to the construction of an interpretation, learning via memorization may occur, but understanding will not. Accordingly, determining whether a student has knowledge relevant to the interpretation of the to-be-experienced instruction or whether the student has knowledge that can be activated in a mnemonic activity or a learning-by-analogy activity seems of considerable importance. If the student has relevant background knowledge, the preconditions for understanding are in place and no special educational intervention is needed. If the student does not have the necessary background knowledge but does have useful knowledge that can be activated using mnemonics or analogical techniques, these techniques can be incorporated in the instructional sequence. In the situation where the student does not have relevant background knowledge or knowledge that can be activated to support mnemonic or analogical activities, the students must be taught something new that can bridge the gap between what they already know and what they are going to learn.

It should be emphasized that there is a distinction between assessing specific content knowledge relevant to the interpretation of instructional material and assessing aptitudes that may be predictive of future performance. This difference relates to the edumetric/psychometric distinction discussed by Carver (1974) and to the different purposes of comprehension assessment discussed by Royer and Cunningham (1981).

One of the major issues facing researchers interested in developing instruction designed to enhance understanding is how to devise techniques to assess knowledge that students already have and that will be relevant to the interpretation and acquisition of subsequently encountered instruction. One promising way of doing this is the Sentence Verification Technique (SVT) of measuring comprehension, developed by Royer. The technique involves constructing four test sentences for every original sentence that appears in a text. The first kind of test sentence is called an original and is the exact repetition of a sentence as it appeared in the passage. The second type of test sentence, called a paraphrase, is developed by changing as many words as possible in an original sentence without altering the meaning. The third type of test sentence is called a meaning change and is developed by changing one or two words in an original sentence in a manner that alters the meaning of the sentence. The fourth kind of test sentence is called a distractor and is a sentence that is related to the theme of the passage but unrelated to any sentence appearing in the passage.

An SVT test is constructed by selecting one of the four types of test sentences to represent each original sentence in the passage. The sentences are combined such that each test sentence type is equally represented. The test is administered by having the examinee read the original passage and then to respond "old" or "new" to each of the test sentences (without reexamining the passage). Old sentences are defined as those that are the same as or mean the same as original sentences (originals and paraphrases), and new sentences are defined as sentences that mean something different from an original sentence (meaning changes and distractors).

The rationale for using the SVT as a measure of comprehension is based on the assumption that when an oral or written message is understood, it is represented in memory in a form that preserves the meaning of what was read or heard but not in the exact words. Hence, subjects who have understood the message should be able to correctly identify original and paraphrase sentences as measuring the same thing as their memory representation, and to correctly reject meaning changes and distractors as meaning something different from their memory representations.

To this point, Royer and his associates have accumulated considerable evidence supporting the construct validity of the SVT as a measure of both written and oral comprehension. Briefly, this evidence shows the following: (a) The SVT is sensitive to text readability (Royer, Hastings, & Hook, 1979; Royer, Kulhavy, Lee, & Peterson, in press); (b) the SVT is sensitive to reading skill as defined by grade level (e.g., 4th versus 6th graders) and by teacher-defined reading groups (Rasool & Royer, in press; Royer et al., 1979; Royer & Hambleton, in press; Royer, Kulhavy, Lee, & Peterson, in press); (c) the SVT is sensitive to differences in subject-matter expertise when experts or non-experts read technical text (Royer, Lynch, Hambleton, & Bulgarelli, in press); (d) the SVT is sensitive to Kintschian text characteristics (Royer, Lynch, Hambleton, & Bulgarelli, in press); (e) the SVT correlates positively with standardized reading comprehension test performance, tests of intelligence, and several other measures of cognitive performance (Royer et al., 1979; Royer, Kulhavy, Lee, & Peterson, in press); (f) the SVT measures both reading and listening comprehension in a theoretically consistent manner (Royer, Kulhavy, Lee, & Peterson, in press); and (g) the SVT is a highly reliable measure of comprehension (Royer & Hambleton, in press; Royer, Kulhavy, Lee, & Peterson, in press).

The possibility of using the SVT as a measure of specific subject-matter knowledge stems from another line of research undertaken by Royer and his associates. This research is based on the previously discussed assumption that relevant prior knowledge is a prerequisite of language comprehension. Thus, the extent to which someone can read and understand a text based on subject matter is an index of the relevant knowledge the person has about the subject matter. Moreover, this leads to the possibility that the amount of knowledge a person has in a particular subject-matter area will be a predictor of the amount of knowledge the person can acquire from a course of study in that content area.

Based on the above line of reasoning, Royer and Marchant have initiated a series of studies in which on the first day of class, students enrolled in a college-level, introductory psychology course read abstracts of psychology journal articles and take SVTs based on those articles. We now have data from around 200 students and have found that correlations between performance on SVTs administered at the beginning of the school year and end-of-course performance was .46. In comparison, the correlations between Scholastic Aptitude Test (SAT) scores and end-of-course performance was .39.

The next step in this research program is to expand the range of tests that are examined. For example, we intend to replicate the predictive studies in courses involving the natural sciences. At some point, we would also like to conduct studies that involve predicting physical performance; for example, measuring comprehension of a set of instructions describing a physical task using the SVT, and then using actual ability to perform the task as a criterion measure.

The SVT procedure for measuring comprehension represents only one way to attempt to increase the predictive capability of instruments used to select personnel for specific training experiences. There are undoubtedly many ways to accomplish this end, and cognitive theory would suggest that any procedure that assesses the extent of prior knowledge that will be relevant to a subsequent training experience would be useful as a prediction of degree of success in training.

An ISD Model Designed to Produce Understanding

The ideas presented thus far in the paper can be incorporated in a revision of the familiar Instructional Systems Development (ISD) model. The revised ISD model I have in mind is presented in Figure 6-2. It should be noted that my revised model represents one approach to developing instruction that has performance validity, as Irv Goldstein has used that term in his paper. I would also like to note that the revised model I have in mind is not meant to replace the traditional ISD model, but instead, to complement it. To use Irv's taxonomy, the traditional ISD model represents one approach to training validity, whereas my revision represents one approach to performance validity. Both are obviously critical to any training program. Moreover, I think many of the ideas present in Tom Shuell's paper are especially relevant to developing instruction that has training validity.

Transfer task analysis and transfer objectives. The model begins with what I call transfer task analysis and the development of transfer objectives. The idea behind transfer task analysis is to analyze the domain of activities a trainee might engage in after training, in an effort to identify types of tasks that are represented in virtually every job activity a trainee might engage in. Those activities can then serve as indices of the ability of training to transfer to relevant after-training activities. Based on what I have read in Bob Gagné's paper, it would appear that considerable work of this type has already been completed.

Once a set of transfer activities has been identified, the next step is to define transfer objectives. Transfer objectives can be thought of as being distributed along a continuum of near transfer to far transfer, with near-transfer tasking involving conditions very similar to those operative during training, and far transfer involving conditions very different from those involved in training (Royer, 1979). As an example, imagine a training sequence directed at the repair of radio receivers. A near-transfer objective might ask the trainee to read a repair manual on a receiver similar to the ones used in training and then take a comprehension test to

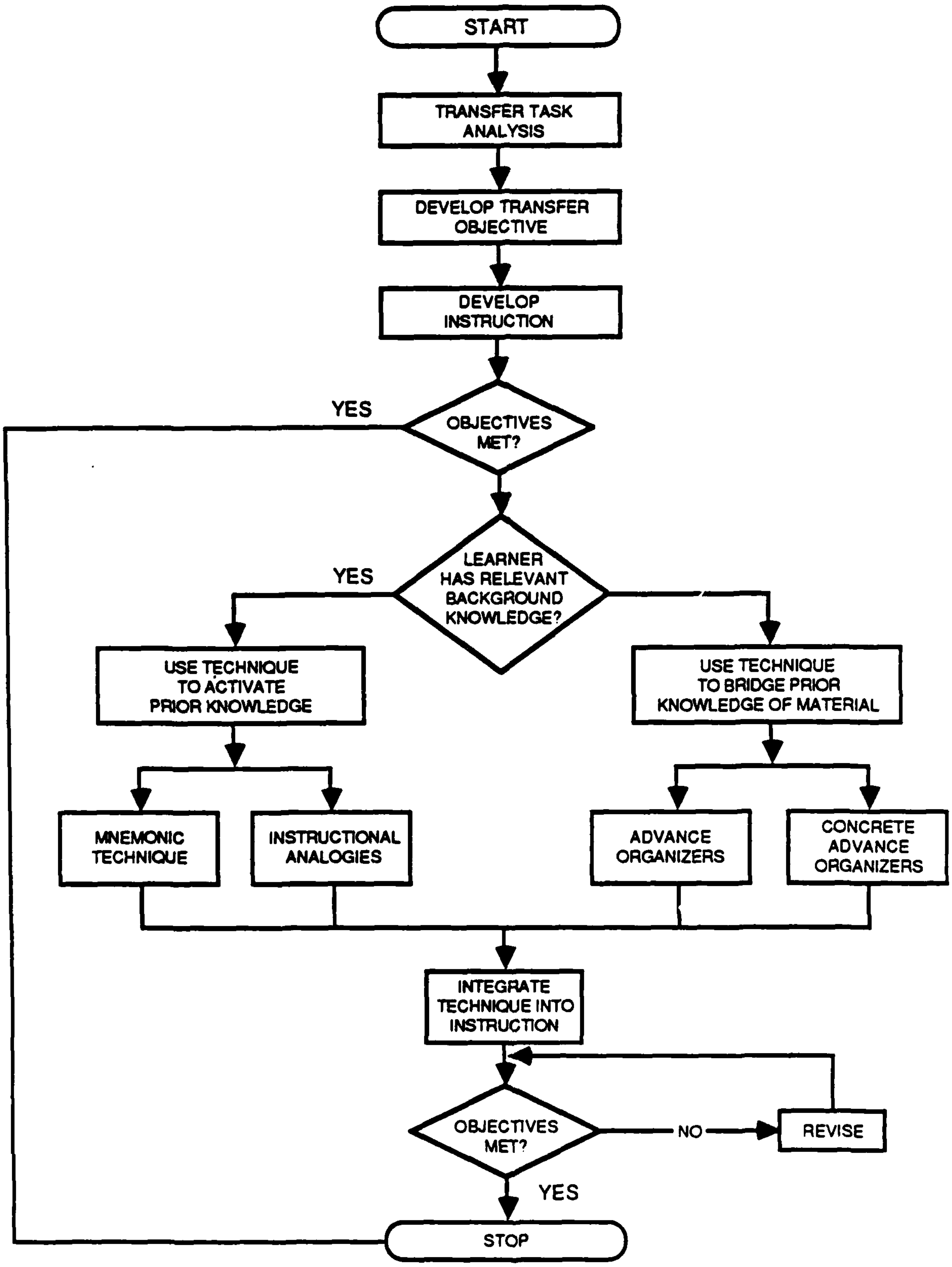


Figure 6-2. Revised ISD Model.

see if what was read was understood; or, as another example, the trainee might be expected to perform a familiar test operation on a receiver that had never been seen before, but was similar to those used in training. An example of a far-transfer task might involve asking a trainee to design and construct a circuit for an inoperable receiver from raw components. This might include the simulation of field conditions in which the trainee would have to exhibit inventiveness in order to satisfy the objective.

It is likely that most transfer objectives are going to fall in the range of near- to mid-transfer level. Satisfactory performance on objectives at this level would, I think, be consistent with what Bob Glaser imagines an expert apprentice to be able to do. There may, however, be situations calling for the capability of performing far-transfer activities. For example, training of special forces units may very well call for mastery of skills falling into the far-transfer range.

Developing and trying out instruction. The next phase of the revised ISD model would entail developing instruction and then trying it out in a small-scale test. Instructional development at this point would not involve any special attempts to enhance understanding. After the instruction was developed, it could be tried out on a small scale to see if the transfer objectives were satisfied. If they were, no further instructional development would be necessary.

Building in techniques to enhance understanding. If performance on the transfer objectives is not satisfactory, the next step in the model is to build-in techniques designed to enhance understanding. The first stage of this process entails determining whether trainees have background knowledge on which to base an approach to developing understanding. In general, this would probably be done intuitively. For example, if the instructional designer has an analogy in mind that might enhance understanding, he or she would also be likely to have a good sense of whether the trainees have the knowledge on which the analogy is based. Alternatively, if there did not appear to be a knowledge base from which an approach to understanding could be developed, a technique to establish a bridge between prior knowledge and the material to be learned would be required.

The knowledge-enhancing technique that was chosen would depend on the nature of the material to be learned. For instance, if the training involved the acquisition of a set of terms or concepts, it would be appropriate to use a mnemonic strategy. Alternatively, if a more complex principle or rule were involved, analogies, advanced organizers, or concrete models might be appropriate.

Integrating understanding treatments into instruction. After an approach to enhancing understanding had been selected and instruction developed, the final stage of the model would entail integrating the new material into the previously developed instruction. This would be followed by another tryout and revision cycle until the transfer objectives had been satisfied.

Research Questions for the Learning Research Laboratory

In my paper I have tried to present what I perceive to be a developing consensus on what understanding is and how it might be operationally defined, measured, and enhanced via instruction. Many of these ideas are tentative and need to be developed further. Below are several programs of research that would contribute to the development process.

Assessment of prior knowledge as a predictor of training performance. The working hypothesis here is that assessment of knowledge specifically relevant to a particular course of instruction will be a better predictor of end-of-training performance than will more general aptitude measures. The key to evaluating this hypothesis is to devise measurement procedures that entail performance of some activity actually involved during training. Examples might be the ability to comprehend text samples from an actual training manual or the assessment of learning from a "mini-lesson" derived from the training sequence.

Transfer objectives as indices of understanding. There are two lines of research I will propose in this area. The first would entail the development of a methodology for developing transfer objectives. Part of this would involve developing a procedure for isolating tasks that are common to a domain of work activity. The next part would entail developing a procedure for developing transfer objectives from the identified tasks common to the domain. This would involve, for example, establishing guidelines for the selection of transfer objectives that range along the near-to-far-transfer dimension.

The second line of research in this area would be the evolution of the hypothesis that trainees who can perform at a satisfactory level on transfer objectives at the end of training will be able to perform better on the job than will trainees who did not attain satisfactory performance on the transfer objectives. In short, there needs to be research on whether transfer objectives can serve as the measurement base for the performance validity of a training sequence.

Research on techniques designed to enhance understanding. Earlier in the paper, four techniques were discussed that either activated relevant prior knowledge or that established a bridge between prior knowledge and material to be learned. Whereas there is much literature establishing the utility of these techniques in a variety of educational situations, I know of no research that examines the utility of the techniques in military training situations. I think three of the techniques are especially promising. Mnemonic techniques similar to the keyword method have been shown to be extremely powerful techniques for learning a variety of materials, and their utility for Air Force training problems should be examined. I think the use of analogies for instructional purposes is just beginning to be examined in detail, and their use holds a great deal of promise for developing instruction that enhances understanding. Concrete models are also very promising. Rich Mayer's research has already shown the effectiveness of such models, and I would anticipate positive effects could also be shown in military training situations. I am less optimistic about advance organizers, even though there is substantial evidence showing their effectiveness. These effects are, however, generally small, and there is still substantial doubt about what an advance organizer is.

REFERENCES

Anderson, R. C. (1972). How to construct achievement tests to measure comprehension. Review of Educational Research, 42, 145-170.

Anderson, R. C., & Biddle, W. B. (1975). On asking people questions about what they are reading. In G. Bower (Ed.), Psychology of learning and motivation (Vol. 9). New York: Academic Press.

Andre, T. (1979). Does answering higher-level questions while reading facilitate productive learning? Review of Educational Research, 49, 280-318.

Atkinson, R. C. (1975). Mnemotechnics in second-language learning. American Psychologist, 30, 821-828.

Atkinson, R. C., & Raugh, M. R. (1975). An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. Journal of Experimental Psychology: Human Learning and Memory, 104, 126-133.

Ausubel, D. P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. Journal of Educational Psychology, 51, 267-272.

Ausubel, D. P. (1963). The psychology of meaningful verbal learning. New York: Grune & Stratton.

Ausubel, D. P., & Fitzgerald, D. (1961). The role of discriminability in meaningful verbal learning and retention. Journal of Educational Psychology, 52, 266-274.

Ausubel, D. P., & Fitzgerald, D. (1962). Organizers, general background, and antecedent learning variables in sequential verbal learning. Journal of Educational Psychology, 53, 243-249.

Ausubel, D. P., Novak, J. D., & Hanesion, H. (1978). Educational psychology: A cognitive view (2nd ed.). New York: Holt, Rinehart, and Winston.

Ausubel, D. P., & Youssef, M. (1963). Role of discriminability in meaningful parallel learning. Journal of Educational Psychology, 54, 331-336.

Bransford, J. D., & McCarrell, N. S. (1974). A sketch of a cognitive approach to comprehension: Some thoughts about understanding what it means to comprehend. In W. B. Weiner & D. S. Palermo (Eds.), Cognition and the symbolic processes. Hillsdale, NJ: Erlbaum.

Carver, R. P. (1974). Two dimensions of tests: Psychometric and edumetric. American Psychologist, 29, 512-518.

Dansereau, D. F. (1978). The development of a learning strategy curriculum. In H. O'Neil (Ed.), Learning strategies. New York: Academic Press.

Dansereau, D. F., Actkinson, T. R., Long, G. L., & McDonald, B. A. (1974, December). Learning strategies: A review and synthesis of the current literature. (AFHRL-TR-74-70, AD-A007 722). Lowry AFB, CO: Technical Training Division, Air Force Human Resources Laboratory.

Dansereau, D. F., Collins, K. W., McDonald, B. A., Holley, C. D., Garland, J., Drekhoff, G., & Evans, S. H. (1979). Development and evaluation of a learning strategy program. Journal of Educational Psychology, 71, 64-73.

Duchastel, P. C., & Merrill, P. F. (1973). The effects of behavioral objectives on learning: A review of empirical studies. Review of Educational Research, 43, 53-70.

Fitzgerald, D., & Ausubel, D. P. (1963). Cognitive versus affective factors in learning and retention of controversial material. Journal of Educational Psychology, 54, 73-84.

Gates, A. I. (1917). Recitation as a factor in memorizing. Archives of Psychology, 6, 1-104.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. Cognitive Psychology, 15, 1-38.

Grotelueschen, A., & Sjogren, D. D. (1968). Effects of differentially structured introductory materials and learning tasks on learning and transfer. American Educational Research Journal, 5, 191-202.

Jenkins, J. J. (1979). Four points to remember: A tetrahedral model and memory experiments. In L. S. Cernak & F. I. M. Craik (Eds.), Levels of processing in human memory. Hillsdale, NJ: Erlbaum.

Kaplan, R. (1976). Effects of grouping and response characteristics of instructional objectives on learning from prose. Journal of Educational Psychology, 68, 424-430.

Kaplan, R., & Rothkopf, E. Z. (1974). Instructional objectives as directions to learners: Effect of passage length and amount of objective relevant content. Journal of Educational Psychology, 66, 448-456.

Kaplan, R., & Simmons, F. G. (1974). Effects of instructional objectives used as orienting stimuli or as summary/review upon prose learning. Journal of Educational Psychology, 66, 614-622.

Levin, J. R., Pressley, M., McCormick, C. B., Miller, G. E., & Shriberg, L. K. (1979). Assessing the classroom potential of the keyword method. Journal of Educational Psychology, 71, 583-594.

Lorayne, H., & Lucas, J. (1974). The memory book. New York: Ballantine.

Mager, R. F. (1962). Preparing instructional objectives. Palo Alto, CA: Feron Publishers.

Mayer, R. E. (1975). Different problem solving competencies established in learning computer programming with and without meaningful models. Journal of Educational Psychology, 65, 725-734.

Mayer, R. E. (1976). Some conditions of meaningful learning of computer programming: Advance organizers sequencing. Journal of Educational Psychology, 68, 143-150.

Mayer, R. E., & Bromage, B. K. (1980). Different recall protocols for technical texts due to advance organizers. Journal of Educational Psychology, 72, 209-225.

Merrill, M. D., & Strolurow, L. M. (1966). Hierarchical preview versus problem oriented review in learning an imaginary science. American Educational Research Journal, 3, 251-261.

Nassif-Royer, P. (1977). Effects of specificity and position of written instructional objectives on learning from lecture. Journal of Educational Psychology, 69, 40-45.

Perkins, M. R. (1978). Measures of cognitive structure: Do they really assess learning at the level of comprehension. Unpublished doctoral dissertation, University of Massachusetts.

Pressley, M., & Levin, J. R. (1978). Developmental constraints associated with children's use of the keyword method of foreign langauge vocabulary learning. Journal of Experimental Child Psychology, 26, 359-372.

Pressley, M., Levin, J. R., & Miller, G. E. (1980, March). How does the keyword method affect vocabulary comprehension and usage? Working paper No. 278. Wisconsin Research and Development Center for Individualized Schooling.

Rasool, J. M., & Royer, J. M. (in press). Reading competence and narrative and descriptive texts: The assessment of comprehension using the Sentence Verification Technique.

Raugh, M. R., & Atkinson, R. C. (1975). A mnemonic method for learning a second-language vocabulary. Journal of Educational Psychology, 67, 1-16.

Reddy, M. (1979). The conduit metaphor - A case of frame conflict in our language about language. In A. Ortony (Ed.), Metaphor and thought. New York: Cambridge University Press.

Rickards, J. P. (1979). Adjunct postquestions in text: A critical review of methods and processes. Review of Educational Research, 49, 181-196.

Rothkopf, E. Z. (1966). Learning from written instructive material: An exploration of the control of inspection behavior by test-like events. American Educational Research Journal, 3, 241-249.

Rothkopf, E. Z. (1968). Two scientific approaches to the management of instruction. In R. M. Gagné & W. J. Gephart (Eds.), Learning research and school subjects. Itasca, IL: F. E. Peacock.

Rothkopf, E. Z., & Kaplan, R. (1972). Exploration of the effect of density and specificity of instructional objectives on learning from text. Journal of Educational Psychology, 63, 295-302.

Royer, J. M. (1979). Theories of the transfer of learning. Educational Psychologist, 14, 53-69.

Royer, J. M., & Allan, R. G. (1978). Psychology of learning: Educational applications. New York: John Wiley.

Royer, J. M., Bates, J. A., & Konold, C. E. (1983). Learning from text: Methods of affecting reader intent. In A. H. Urquhart & J. C. Alderson (Eds.), Reading in a second language. London: Longman.

Royer, J. M., & Cable, G. W. (1975). Facilitated learning in connected discourse. Journal of Educational Psychology, 67, 116-123.

Royer, J. M., & Cable, G. W. (1976). Illustrations, analogies, and facilitative transfer in prose learning. Journal of Educational Psychology, 68, 205-209.

Royer, J. M., & Cunningham, D. J. (1981). On the theory and measurement of reading comprehension. Contemporary Educational Psychology, 6, 187-216.

Royer, J. M., & Hambleton, R. K. (in press). The reliability of the Sentence Verification Technique as a measure of reading comprehension.

Royer, J. M., Hastings, C. N., & Hook, C. (1979). A sentence verification technique for measuring reading comprehension. Journal of Reading Behavior, 11, 355-363,

Royer, J. M., Kulhavy, R. W., Lee, J. B., & Peterson, S. E. (in press). The Sentence Verification Technique as a measure of listening and reading comprehension.

Royer, J. M., Lynch, D. J., Hambleton, R. K., & Bulgarelli, C. (in press). Using the Sentence Verification Technique to assess the comprehension of technical text by experts and non-experts.

Royer, J.M., & Marchant, H. (in preparation). Predicting college course performance by administering reading comprehension tests at the beginning of the semester.

Royer, J. M., & Perkins, M. R. (1977). Facilitative transfer in prose learning over an extended time period. Journal of Reading Behavior, 9, 185-188.

Scandura, J. M., & Wells, J. N. (1967). Advance organizers in learning abstract mathematics. American Educational Research Journal, 4, 295-301.

Watts, G. H., & Anderson, R. C. (1971). Effects of three types of inserted questions on learning from prose. Journal of Educational Psychology, 62, 387-394.

CHAPTER 7

THE RELATIONSHIP OF TRAINING GOALS TO BASIC RESEARCH ISSUES IN INSTRUCTION

Irwin L. Goldstein
University of Maryland

Instructional systems philosophy as developed by Gagné and his colleagues (e.g., Briggs, Campeau, Gagné, & May, 1967; Gagné, 1970) has had a profound impact on the way that such systems are conceptualized. Similar developments stimulated by these instructional theorists have affected the design of training systems in both the public and private sectors. These approaches (e.g., Goldstein, 1974, 1978) typically emphasize the specification of instructional objectives based on needs assessment procedures, precisely controlled learning experiences to achieve these objectives, criteria for performance, and evaluation information. In a training framework, it is possible to examine the components of instructional systems models and ask what the user desires to achieve; i.e., "What are the goals of the instructional system?" An examination of those goals and of the supporting instructional system components necessary to achieve them suggests a number of areas that require further research and study. This paper examines these goals, the components of instructional systems, and some suggested research issues.

Goals of Instructional Programs

As presented in Figure 7-1, there are four potential goals for training systems:

1. Training validity. This particular goal refers only to the validity of the training program. Validity is determined by the performance of trainees on criteria established for the training program.

2. Performance validity. This goal refers to the validity of the training program as measured by performance in the transfer of what has been learned to the job setting.

3. Intra-organizational validity. This concept refers to the performance of a new group of trainees within the same organization for which the original training program was developed. In this instance, the question is whether generalization will occur to the performance of new trainees based on the evaluated performance of a previous group.

4. Inter-organizational validity. In this instance, the goal is to determine whether a training program validated in one organization can be utilized in another organization.

A consideration of these categories indicates that the achievement of validity at each succeeding stage is affected by an increasing number of variables. For example, the establishment of training validity requires the consideration of evaluation procedures, including criteria of training success. Trainee performance on the job (performance validity) not only requires the consideration of training criteria of success but also demands an examination of the potential disruptive effects of organizational constraints. For example, the training program and the organization often specify different performance objectives, and trainees are caught in a conflict which sometimes results in failure. As a result, the trainee and training program may be declared inadequate because the training analyst has not considered the relevant variables which determine success or failure at each of the four stages of validity. A consideration of

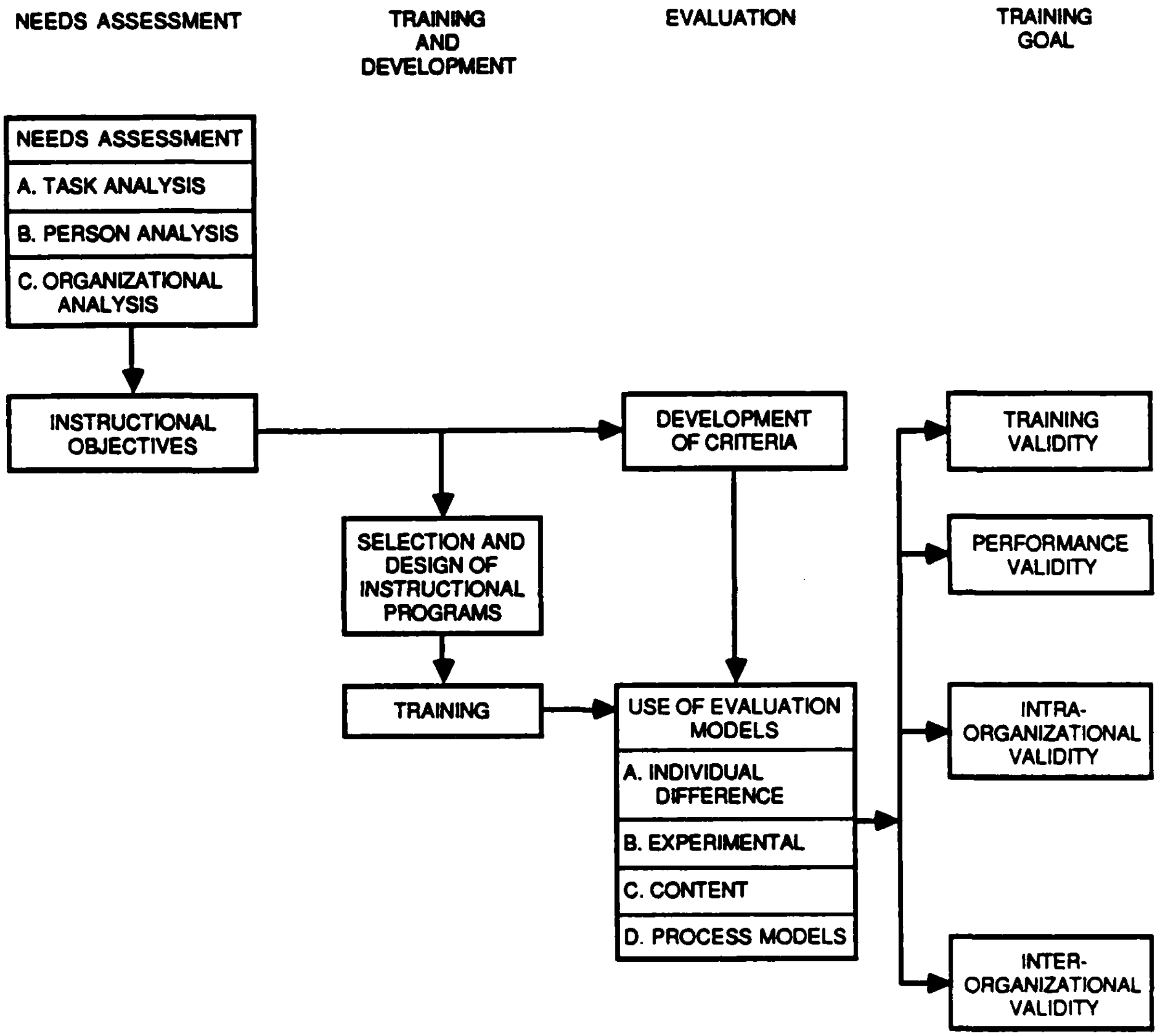


Figure 7-1. The Relationship of Training System Goals and Instructional Design Components.

each of these goals and their related instructional processes is presented next. After a discussion of these issues, some research program issues are suggested.

Training Validity

As noted above, training validity refers to the establishment of treatment effects as a result of the instructional program. In this situation, the pursuit of validity refers only to performance in the instructional program and does not consider transfer performance. While training analysts might note that what occurs in the training program does not provide any guarantees about learning transfer to the job, it should be obvious that the vast majority of instructional programs attend only to training performance and not to transfer performance. Of course, many would argue that even when training programs are not designed on the basis of a needs assessment of the job, there is still some mental appreciation of the transfer performance around which the instructional program is designed. For example, there are many managerial training programs that have as a goal instructing trainees in the basic theory of management. Evidently, managers are expected to return to their array of jobs and situations with what they have learned and are to determine a way to apply it. If one questions the developers of such programs, they will readily admit that they have not done a needs assessment and have not measured transfer performance, but somehow or other they have a "feel" for the job setting and thus have designed their program appropriately. Evidence for this viewpoint is nonexistent. It is necessary to simply accept the fact that this type of training program is concerned only with training performance; it is focusing only on the establishment of training validity. As we shall see, that task in itself is not simple.

It is interesting to note that some programs are primarily designed to achieve training validity. Thus, instructional programs in academic environments are often designed to teach a particular subject matter primarily because of the value of the content and not necessarily because the learning is expected to transfer to other specific settings. Also, the purpose of the implementation of some training programs is to provide a research vehicle. For example, an investigator might be concerned with the effects of different types of feedback on the performance of trainees learning a complex motor skill. In this case, the program is concerned primarily with training validity, not necessarily performance validity. It is possible to debate whether or not training programs should be permitted to focus only on training validity without a concern for the transfer setting. However, that value judgement is not the focal point of this discussion. Rather, the point here is that such programs do exist and that the establishment of training validity in itself requires the consideration of many important variables. Some of the most important points to consider include the following.

Needs Assessment for Training Validity

First, it is important to realize that some form of needs assessment is necessary. In the case of establishing training validity, the needs assessment would not necessarily focus on the behaviors required in the transfer setting. Rather, the needs assessment would attend to the particular goals and objectives to be achieved as a result of the instructional program. Thus, if the program is designed to provide a working knowledge of management philosophy, it is necessary to determine through task analysis and specification of instructional objectives what particular goals the training program is trying to achieve. It is also important to perform a person analysis so that the instructional program is designed to fit the knowledge, skill and ability characteristics of the trainee population. Indeed, this needs assessment must provide the input for the design of the instructional sequences, as well as the suggested criteria by which the training program will be evaluated. It is not reasonable to skip the needs assessment process just because what is being planned is a generalized training program that is not geared to specific behaviors to be performed in a transfer setting. There should still be performance objectives, designed for particular trainee populations, that are determined by needs assessment procedures. Unfortunately, instead of careful needs assessment, the training field often appears dominated by a fads approach. Analysts seldom find out very much about their particular approach before they are off examining another type of program. The fads approach places a heavy emphasis on the development of techniques, without consideration of needs assessment followed by a matching of the techniques to the needs. Training analysts still have to be warned about selecting tools and finding something they fit, by quotes like the following:

> If you don't have a gadget called a teaching machine, don't get one. Don't buy one; don't borrow one; don't steal one. If you have such a gadget, get rid of it. Don't give it away, for someone else might use it. This is a most practical rule based on empirical facts from considerable observation. If you begin with a device of any kind, you will try to develop the teaching program to fit that device. (Gilbert, 1960, p. 478).

Gilbert is not saying that teaching machines, or sensitivity training, or computer-assisted instruction (CAI), or any other technique does not work. He is saying that the design of change programs cannot begin with instructional media. Instead, one must, through needs assessment, determine the objectives of training so that the criteria for evaluation and the choice of instructional program are based on sound decisions. It appears to be unfortunate but true that programs that are designed to achieve only training validity are also more likely to be designed on the basis of the most popular fad rather than careful needs assessment.

Research Issues for Needs Assessment and Training Validity

Interestingly, even the apparently simple issues concerning needs assessment for training validity raise some very complex issues. Training analysts remain uncertain about what the appropriate input should be for the design of instructional programs. Needs assessment of jobs has focused on a number of different factors including the determination of the tasks to be performed; the knowledge, skills and abilities necessary to perform the tasks; and the behaviors necessary to demonstrate competent performance. However, there is virtually no information available comparing the various systems as inputs to the design of training systems. There is not even much information on the best systems to determine which characteristics (e.g., frequency, criticality) should be measured to provide information about: tasks, knowledge, skills or abilities, or performance. The opportunity to perform needs assessment on relevant jobs and to examine the effects of the system utilized, on instructional design and learner performance, is a unique opportunity for a basic learning laboratory.

Criteria and Training Validity

In general, evaluation, as described by Goldstein and Buxton (1982), is considered to be an information-gathering process that should not be expected to lead to decisions that declare a program as all good or all poor. One of the purposes of the needs assessment is the determination of the knowledge, skills, and abilities (KSAs) required for successful job performance. As shown in the model in Figure 7-1, that information must provide direct input to the training program to determine the actual content of the instructional material. The same information concerning the KSAs necessary for successful job performance should also provide the input for the establishment of measures of training success. Logically, we should want our training program to consist of the materials necessary to develop the KSAs to perform successfully on the job. Just as logically, we should determine the success of our training program by developing measures (or criteria) that tell the training evaluator how well the training program does in teaching the trainees the same KSAs necessary for job success. Also, these criteria are used at the end of the training program to determine how well our program is doing at that time; and as discussed in the section on transfer validity, these criteria should also be used later on, when the trainees are on the job, to determine how much of the KSAs learned in training transferred to the actual job.

Research Issues Concerning Criteria for Training Validity

In general, investigations of the types of criteria employed to provide evaluation data are not reassuring. For example, Catalanello and Kirkpatrick (1968) found that of 154 companies surveyed, the largest number (77%) stressed studies related to reports of personal reactions. Even in those instances where such data are collected, some investigators (Mindak & Anderson, 1971) have suggested that most of these measures were "eyeball" attempts to measure reactions. Few studies have bothered to measure learning and on-the-job behavior. It is probably not unreasonable to suspect that these investigations did not consider criterion relevance and reliability. The question about how criterion relevance is determined is not easily answered. There should be a tracing-out process from the needs assessment to the development of criteria. The rules of this process are unknown. Indeed, no one is really very sure what the input should be to maximize criterion relevance. Recently, there has been increased attention to criterion-referenced measures employing an absolute standard of quality (Swezey, 1978), rather than norm-referenced measures that are dependent on a relative standard. Criterion-referenced measures provide a standard of achievement for the individual based on specific behavior objectives and thereby provide an indication of the degree of competence attained by the trainee. The characteristics of criterion-referenced measures - including methods of determining content validity as well as the design of methods to determine reliability - need attention.

There is also great difficulty in determining what constitutes ideal performance and how it should be measured. Thus, Williges, Roscoe, and Williges (1972) note that there is little agreement on what constitutes ideal pilot performance, and the reliability of most pilot-performance grading systems has been disappointing. Consequently, even studies of original learning are difficult to complete because there is little agreement on what constitutes terminal performance. Even in the area of simulators, where there have been extensive efforts, the establishment of appropriate criteria for complex performance is a serious issue demanding the attention of researchers. In addition, there are continuing questions concerning the development of new criterion measures that provide further information about the instruction process. Thus, reaction measures which only ask the trainee for his general feelings about learning have not been found to be particularly useful. However, it is possible that trainees are able to self-assess their learning of particular components of KSAs. Research is needed to determine whether learners can assess their own capabilities concerning how much they have learned. There are also questions concerning the relationships between these self-assessment indices and other measures of learning in the instructional environment.

Evaluation Models and Training Validity

Instructional programs are never complete; instead, they are designed to be revised on the basis of information obtained from evaluations founded on relevant multiple criteria that are reliable and free from contamination. The better experimental procedures control more of the potentially disruptive variables, thereby permitting a greater degree of confidence in specifying program effects. Although the constraints of the training environment may make laboratory-type evaluations impossible to achieve, an awareness of the important factors in experimental design makes it possible to conduct useful evaluations even under adverse conditions.

There are several types of evaluation in models employed to assess learning. Perhaps, the best-known model is typically titled an experimental or a quasi-experimental design model. These types of models, which are sometimes employed in evaluating training validity, include considerations related to pre- and post-testing, control groups, random assignment of subjects, etc. The job of the analyst is to choose the most rigorous design possible and to be aware of its limitations. Unfortunately, there are few studies that employ the most rigorous design within the limitations imposed by the environment. In many instances, the authors of research articles apologize in advance for poor design that obviously could have been strengthened by some initial effort. It appears that most efforts at evaluation are afterthoughts. Appropriate planning and the use of elementary research procedures would significantly increase the amount of valid information gathered.

Recently, however, training researchers have begun to consider other types of evaluation models. One that is now being used quite frequently is an individual difference model. In this model, learning scores are correlated with later, on-the-job performance scores to determine whether the same pattern of individual differences exists. Thus, if persons who perform well in training also perform well on the job, positive correlations would result. The other model is a content validity model where the analysis examines whether the KSAs that are judged critical for job performance are also those emphasized in training. Further discussion of these models will be presented in the discussion on transfer performance since they all depend on transfer data as well as on training data. They are presented here only because they are also dependent on training validity data. This also is an appropriate place to hint that issues concerning the interrelationship among these different types of evaluation models pose important research questions.

Still another evaluation model in training validity is called a process model. This model is actually a partner to all evaluation models, rather than a completely separate way of examining the world of training validity. There is a growing awareness that the use of evaluation designs has not provided the degree of understanding that was originally anticipated. The evaluation designs and specification of outcome criteria have been based on a product or outcome view of training validity. Thus, researchers collected pre- and post-measures, compared them with control groups and discovered that they did not understand the results they had obtained. This problem was especially apparent when the collectors of these data were outside consultants who appeared only to collect pre- and post-data but had no conception of the process that had occurred in training between the pre- and post-measurement. An event experienced by this author illustrates this issue. In a study of computer-assisted instruction in a school setting (Goldstein & Rosenberg, 1977), two teachers each agreed to instruct a geometry class by traditional methodology and by CAI. Thus, each teacher taught one traditional and one CAI class. Further, the teachers agreed to work together to design an exam that would cover material presented in both classes. At the end of the first testing period, the traditional classes taught by each teacher significantly outperformed the CAI groups taught by these same teachers. The same outcome occurred at the second testing. However, at the third testing, one of the CAI groups improved so that it was now equivalent to the two traditional groups. The other CAI group performed significantly worse than the other three instructional groups. One reasonable conclusion for this series of events is that one of the teachers finally learned how to instruct the CAI group so that it was now equivalent to the two traditional groups, whereas the other teacher had not been able to perform that task with his CAI group. Indeed, if the investigators had collected only the outcome measures, any number of similar erroneous conclusions could probably have been offered as explanations for the data. In this case, the investigators also observed the instructional process, to provide information about the program. The evaluators had learned that the instructor for the CAI group that eventually improved had become disturbed over the performance of his students. Thus, the teacher offered remedial tutoring and essentially turned his CAI class into a traditional group.

Many researchers have become concerned about the lack of understanding about the variables in the training process which affect or determine outcomes. The view is especially well expressed by Cronbach (1973):

> Insofar as possible, evaluation should be used to understand how the course produces its effects and what parameters influence its effectiveness. It is important to learn, for example, that the outcome of a programmed instruction depends very much upon the attitude of the teacher; indeed, this may be more important than to learn that on the average such instruction produces slightly better or worse results than conventional programs. (p. 675)

An understanding of these process issues - which are the ones on which to collect data - and how they interact with learner performance are all important research issues that deserve further attention.

Performance Validity

As noted earlier, the establishment of performance validity has the additional burden of determining whether performance has positively transferred from the training program to the on-the-job environment. For this type of validity, transfer of performance is considered only for the original group being trained in the instructional program. The question of generalizing to new training populations or new organizations will be treated later in this paper. Considering the points presented in the discussion of training validity, it may appear that there are relatively few additional issues to be added to the analysis of performance validity. Unfortunately, that is not the truth. Now it is necessary to consider conditions which involve the transfer of performance in one environment (training) to another environment (on the job). This perspective has important implications for the needs assessment process, because now the objectives and goals of the instructional program must be determined by a needs assessment of the transfer environment.

Needs Assessment and Performance Validity

In the preceding section on training validity, it was noted that the needs assessment process was still necessary. However, it was limited to task and person analysis to establish learning objectives that were not necessarily related to transfer performance. In the establishment of performance validity, the concern is with designing training programs based on a needs assessment of on-the-job performance.

There are dozens of studies that can be cited to demonstrate the dangers of designing training programs without this type of needs assessment. Perhaps one of the sadder examples is revealed by Miller and Zeller (1967) in an investigation of 418 hard-core unemployed trainees in a program to train highway construction machinery operators. The authors were able to obtain information from 270 graduates. Of this group, 61% of the graduates were employed and 39% were unemployed at the time of the interview. In addition, more than half of the employed group said they were without jobs more than 60% of the time. Some of the reasons for the unemployment situation were inadequacies in training, such as not enough task practice. The details showed that the program was not based on a consideration of job components. One trainee noted that "contractors laughed when I showed them my training diploma and said come back after you get some schooling, buddy" (p. 32-33). In a familiar lament, the authors of the report wonder how a training program could be designed without a thorough analysis of the tasks and the KSAs required.

As far as research needs concerning traditional needs assessment, they are all specified in the previous section on training validity. However, there is a growing realization that training occurs in one organization (training organization) and performance occurs in another organization (the work organization). Thus, performance validity forces consideration of the fact that something learned in one environment (training) will be performed in another environment (on the job). The trainee will enter a new environment, to be affected by all of the interacting components that represent organizations today. Certainly, there are some aspects of the environment which help determine the success or failure of training programs beyond the attributes the trainee must gain as a result of attending the instructional program.

The actual components of an organizational analysis are dependent on the type of program being instituted and the characteristics of the organization to which transfer will occur. However, the following broad categories should be considered.

1. The Specification of Goals

When organizational goals are not considered in the implementation of training programs, objectives and criteria that ensue from the needs assessment process are not evaluated. Later, the organizations are not able to specify their achievements because they have not collected the necessary information. Clearly, if the goals are not specified, it is not very likely that they are included in the design of the instructional program.

Organizations might expect training programs to provide trainees with expectations about the job, or particular views toward performance requirements. It is not unusual, for example, to discover that police training programs are devoted to skill requirements (e.g., operating a police vehicle or utilizing firearms) or information requirements (knowing the difference between a felony and misdemeanor). Yet, organizational analyses often discover that there are organizational expectations concerning interpersonal relationships with the public and a concern for all citizens regardless of race, color, or creed. Obviously, if these organizational philosophies and emphases are not clearly defined and specified during the needs assessment phase, they will not be considered in the design of the training program and, at best, will receive only passing attention in the instructional sequence. In this case, the problem is further complicated by the fact that it is a lot easier to teach weapons procedure than to instruct in the topic of interpersonal relations. Thus, certain topics may be de-emphasized and sometimes organizations become aware that they have a problem only when they are facing a series of complaints. At that point in time, everyone wonders why the organizational objectives were not translated into training and job requirements. In short, many training programs are based on teaching skills requirements rather than on unspecified and unidentified organizational objectives. Yet, paradoxically, organizations often judge the value of a program on the basis of their own objectives that were never specified, never considered in the design of the program, and never utilized in designing the evaluation model. Thus, the financial cost of some instructional programs results in their early demise because the instructional system analysis did not identify that variable as an organizational objective.

2. The Determination of the Organizational Training Climate

As complex as the specification of system-wide organizational objectives appear to be, the determination of the objectives by themselves will not do the job. Unfortunately, many of the situations are also marked by organizational conflicts, which are very disruptive. For example, conflicts between Government sponsors of a program, employers, and training institutions can completely disrupt programs (Goodman, 1969). Many of these conflicts are based on the different parties to the program having different sets of equally unspecified goals and expectations. A study by Salinger (1973) characterized the negative consequences for a poor climate. In the study, top management in Federal agencies had a generally negative view of training and no direct knowledge of its benefits. This resulted in a system that failed to reward managers for effective training efforts. Rather, managers failed to plan or budget for training. It is not difficult to guess that whatever training was provided did not serve much of a purpose for the individual or the organization. Even more likely, there are insidious conflicts that somehow (with a magic wand) the training program is supposed to resolve. Even assuming that the training program could be appropriately designed in such a situation, it is likely it will succumb to the conflict.

3. The Identification of Relevant External System Factors

The preceding sections on organizational analysis have identified issues related to the failure to specify organizational goals and the problems of organizational conflicts. The issue examined in this section is the failure to recognize the importance of the interacting constraints acting on an organization and their effects upon training programs. These considerations could be treated as a failure to specify organizational goals or as organizational conflicts. However, *external constraints* are becoming a very serious problem in the design of training programs. Thus, they are treated as a separate section. The design of instructional programs is affected by legal, social, economic, and political factors. Interrelationships of these variables should be carefully specified during organizational analysis. An example of these factors is provided by Salinger's study (1973), which found instances of Government clerical workers being taught writing and typing styles that were prohibited on the job. Another example is that Federal requirements concerning equal employment opportunity often affect the applicant pool, which in turn affects the design of instructional programs.

Research Issues Involving Needs Assessment and Performance Validity

In addition to the research issues concerning needs assessment discussed in the training validity section, the organizational analysis issues involved in transfer suggest a number of important research questions. First, it is important to identify organizational constraints to transfer performance. Then, it is necessary to ask how that should affect the design of instructional research programs. For example, it is clear that in some cases supervisors are not ready to reward the behaviors that trainees have gained as a result of instruction, nor are they willing to serve as role models. In that case, one question is "How does that affect the maintenance of learned behavior?" Also, it is possible to build instructional systems that compensate for that problem. All of these kinds of questions could be addressed in a basic learning laboratory.

Evaluation and Performance Validity

A reexamination of the evaluation section for training validity should reinforce the point that each type of validity being considered builds on the concepts presented previously. Thus, performance validity assumes that the training analyst has already met the issues of criterion development and evaluation design appropriate for the establishment of training validity. Clearly, if the concern is performance validity, there is increasing complexity in the design of proper evaluation methodologies. The additional difficulties are well documented by Blaiwes and Regan's (1970) suggestion that evaluation efforts must consider three criteria: (a) original learning efficiency, (b) transfer of learning to the new task, and (c) retention of original learning. Yet, in the area of flight simulation, where the greatest effort has been expended, the emphasis has been on the most immediate criterion, original learning; and even those studies have been plagued by serious problems. Basic research involving development of criteria for both training and performance validity is a necessity.

The consideration of performance validity also features the discussion of two other evaluation models briefly mentioned in the training validity section. They are individual difference models and content validity models.

1. Individual Difference Models

Many psychologists have emphasized the use of training scores as a way to predict the future success of potential employees. There are a number of these types of studies showing meaningful relationships between training performance and on-the-job performance. For example, Kraut (1975) found that peer ratings obtained from managers attending a month-long training course predicted several criteria, including future promotion and performance appraisal ratings of job performance. Other investigators have used early training performance to predict performance later on in more advanced training. An example of this approach is offered by Gordon and Cohen (1973), whose study involved a welding program that was part of a larger manpower development project. The program consisted of 14 different tasks that fell into four categories and ranged in difficulty from simple to complex. Advancement from one task to the next was dependent on successful completion of all previous tasks. Thus, trainees progressed at a rate commensurate with their ability to master the material to be learned. These investigators understood that they were predicting the performance of individuals on a later task (e.g., on the job or later in training) based on performance in the training program. As a matter of fact, once these relationships have been established in an appropriately designed study, it is possible to select individuals for a job or for later training based on these training scores. In other words, the training score serves as a validated predictor of future performance. However, an important note of caution should be considered when this technique is contemplated as an evaluation of the training program. The relationship between training performance and on-the-job performance simply means that persons who perform best on the training test also perform well on the job. This does not necessarily mean that the training program is properly designed or that persons have learned enough in training to perform well on the job. It is entirely possible that the training program did not teach anything or that the trainees did not learn anything. In those cases, there could still be individual differences on the training test. Even if the training program did not teach relevant material for the job, there would still be a strong relationship if the person who did well on the test did well on the job, and the persons who did poorly on the test did poorly on the job. In other words, the training program had not achieved anything, except to maintain the individual differences between trainees that might have existed before they entered training.

2. Content Valid Models

If the needs assessment is appropriately carried out, and the training program is designed to reflect the KSAs, then the program would be judged as having content validity. In other words, the training program should reflect the domain of KSAs represented on the job that the analyst has determined should be learned in the training program. One way of conceptualizing the content validity of a training program is presented in Figure 7-2. In this figure, the horizontal axis across the top of the figure represents the dimension of the importance or criticality of the KSAs as determined by the needs assessment. While the diagram presents KSAs as being only important or not important, it should be realized that this is an oversimplification of a dimension with many points. The vertical dimension represents the degree of emphasis given the KSAs in training. Again, to simplify the presentation, the dimension is presented as indicating that the KSA is or is not emphasized in the training program. This results in the fourfold table presented in Figure 7-2. Using this approach, both boxes A and D provide support for the content validity of the training program. KSAs that fall into box D are judged as being important for the job and should be emphasized in training. Items in box A are judged as not important for the job and should not be emphasized in training. Conceptually, the degree to which KSAs fall into categories A and D make it possible to think about the training program as being content valid. Of course, this is an oversimplification. There will be KSAs that are judged as moderately important for the job or KSAs that are moderately emphasized in training. However, it is possible to conceive of this type of relationship and actually measure the degree to which those KSAs judged as important are emphasized (and hopefully learned) in training. Very few persons would be unhappy with a training program that has a tendency to emphasize the objectives associated with KSAs that are judged critical or important for job performance.

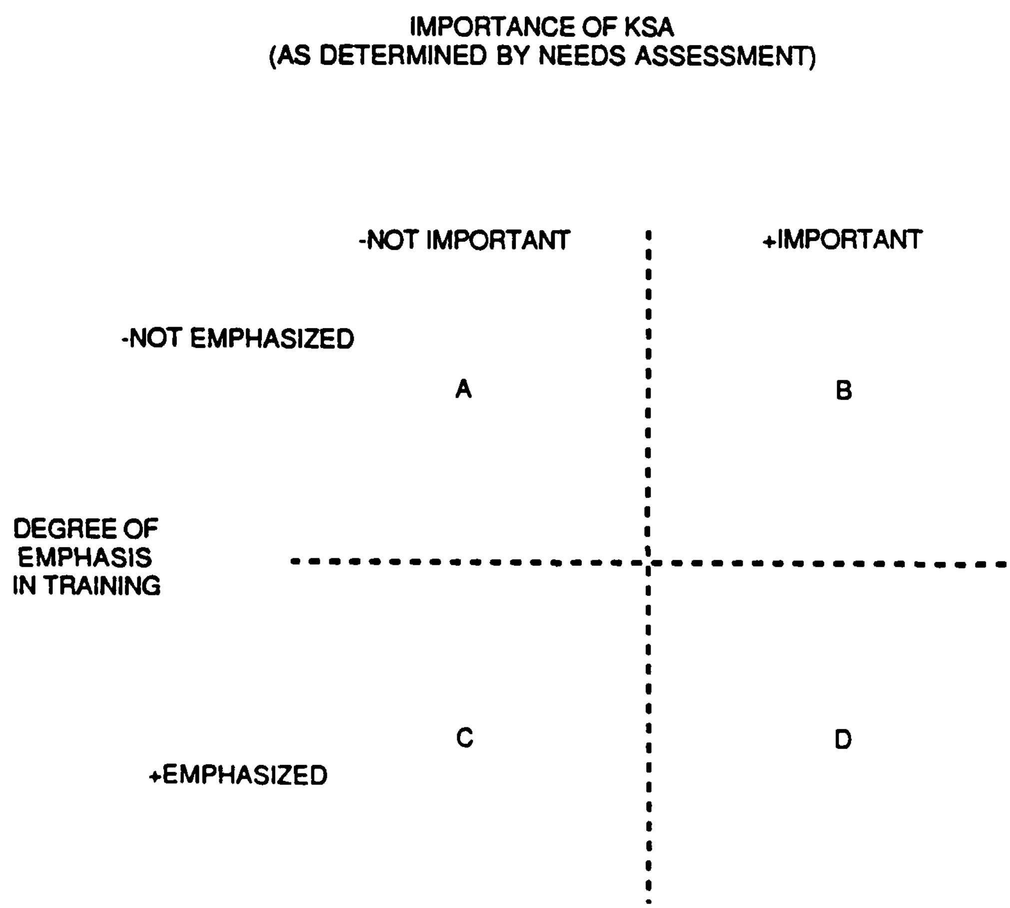


Figure 7-2. A Conceptual Diagram of Content Validity of Training Programs.

Box B represents a potential error and could affect the degree to which a program is judged content valid. KSAs falling into the B-category are judged as important for the job but are not emphasized in training. From a systems perspective, these items must be analyzed to determine whether the organization intends for these KSAs to be gained as a result of training. If that is the case, then there is a problem. However, it is also possible that individuals are expected to be selected with that particular KSA or that the individual is expected to learn the material related to that KSA on the job. To that extent, the training program should not be expected to emphasize the item, and its content validity would not be questioned. However, it would still be important to determine that KSAs judged as important or critical are covered in another system such as selection. If the item is not covered, then the organization must decide whether revision of the training program is necessary or whether some other system has to be redesigned to cover that material.

Box C represents KSAs that are emphasized in training but are judged as not being important for the job. This is often a criticism of training programs. This is, they tend to spend a lot of time emphasizing material that is not job related. Most analysts agree that the use of needs assessment procedures results in a decrease in the amount of training time necessary to complete the program. This usually comes about because of a reduction in the scope of the program, based upon an elimination of the type of items present in category C. Interestingly, a systems view of this process might even suggest a reduction of items that would appear in category D. Selection experts might suggest that sometimes KSAs included in category D are unnecessary because they have already been used as a basis for selection. Thus, trainees are again subjected to training materials on KSAs that are already in their repertoire.

Research Issues Concerning Evaluation and Performance Validity

Besides the criterion issues mentioned earlier, basic research concerning evaluation models is badly needed. There is virtually no research comparing experimental design, individual differences, and content evaluation models. There are very limited data available relating what type of individual difference measure best predicts performance at various stages of learning on different tasks. Also, there are virtually no data available on appropriate measures of degree of content validity nor on the relationship between data on content validity and other evaluation models.

Intra-Organizational Validity

Intra-organizational validity assumes that the trainer has established training and performance validity and is now concerned with predicting the performance of a new group of trainees. Just as performance validity assumes the consideration of the points established for training validity, intra-organizational validity assumes that the points discussed for training and performance validity have been established. It should be noted that in a previously expressed view, evaluation is considered an information-gathering process that provides feedback about the multiple objectives of most training programs. Thus, it becomes apparent that evaluation should be a continual process which provides data as a basis for revisions of the program. New data should be collected based on the performance of a new group of trainees, which provides further understanding about the achievement of objectives or about the variables that affect the achievement of objectives. That does not mean that each effort must start from the very beginning. However, it should be possible to collect further new information about the effects of revisions. Also, it should be possible to collect data that can be checked against previously collected information to make sure that the instructional program is having the same effect.

Research Needs Concerning Intra-Organizational Validity

Although there is a limited amount of data that help us to understand some aspects of training and performance validity, there are virtually no data on intra-organizational validity. There is information that jobs change, persons entering jobs change, and training programs change; but there are no data on how this affects the performance of individuals who enter training programs. Most everyone would agree that it is dangerous to generalize to new trainee populations when they are not being instructed in the same training program. However, most training analysts fail to realize how different established training programs tend to be from the original training program that was evaluated. Some of these difficulties can be labeled problems of reactivity. The excitement generated by observers, experimental equipment, and questions being asked often results in changes in behavior that are not a result of the actual training program. To the extent that these variables are a source of the training result, and to the extent that these variables are not present the next time training is offered, generalization is difficult. There is a certain aura of excitement surrounding new instructional treatments or training programs that simply disappears over time. In some studies (e.g., Rosenthal, 1966), researchers have specified how the expectations of the experimenters (or trainers or teachers) themselves have an effect on the performance of individuals. To the extent that these factors change over time, the training program has changed, and it is difficult to generalize. One of the most compelling reasons for the use of process criteria is to attempt to specify these variables and their effects. When the process criteria indicate that the training programs have not changed, then it is obviously safer to generalize. A training laboratory that uses programs over long periods of time, with different groups of trainees, has a unique opportunity to study these variables. Researchers might even be able to gather data on how programs tend to change. This author will never forget the look of astonishment on the faces of a number of high-level executives who had just discovered that the reason grocery clerks could not operate the cash register was because this instructional sequence was no longer part of the training program.

Inter-Organizational Validity

In this instance, the analyst is attempting to determine whether a training program validated in one organization can be utilized in another organization. All of the factors discussed in training validity, performance validity, and intra-organizational validity affect this decision. As indicated in the section on intra-organizational validity, when the needs assessment shows differences (i.e., the task, person, or organizational components) or the evaluation is questionable or the training program has changed, then generalization is dangerous. In this instance, the needs assessment and evaluation have not been performed for the organization that desires to use the training program. Considering the vast number of ways organizations differ, it is dangerous to attempt to generalize the training results in one organization to trainees in another organization.

Research Needs Concerning Inter-Organizational Validity

One reason it is dangerous to generalize is that we know so little about the effects of variables such as organizational differences on the design, conduct and evaluation of training programs. On the other hand, it is entirely appropriate to borrow needs assessment methodologies, evaluation strategies, and training techniques to try out in a research laboratory. Through these procedures, organizations can establish techniques that work, and perhaps it will be possible to begin to understand what variables affect the success of programs across organizations.

Summary of Research Issues

1. Needs Assessment Issues

Training analysts remain uncertain about what the appropriate input should be for the design of instructional programs. Needs assessment of jobs has focused on a number of different factors including the determination of the tasks to be performed, the KSAs necessary to perform the tasks, and the behaviors necessary to demonstrate competent performance. However, there is virtually no information available comparing the various systems as inputs to the design of training systems. There is not much information on the best systems to determine which characteristics (e.g., frequency, criticality) should be measured to provide information about: tasks, KSAs, or performance. A consideration of the kinds of parameters that are directly relevant to the Air Force might suggest different types of strategies for various KSAs. For example, it may be possible to distinguish between skills and abilities that are frequently performed as compared to those skills and abilities that are critical to performance but are rarely performed. A related issue is that the Comprehensive Occupational Data Analysis Programs (CODAP) task analysis system currently employed by the Air Force includes task-difficulty ratings that are obtained by asking supervisors "How long does it take to learn the task?" A basic learning laboratory could obtain important validation information by comparing the measure with actual learning time on the task. A direct measure of the time it takes to learn a job component would make possible interesting and valuable comparisons with task difficulty as currently assessed. The opportunity to perform needs assessment on relevant jobs and to examine the effects of the utilized system on instructional design and learner performance is a unique opportunity for a basic learning laboratory.

In addition, organizational analysis issues involved in transfer suggest a number of important research questions. First, it is important to identify organizational constraints to transfer performance. Then, it is necessary to ask how that should affect the design of instructional research programs. For example, one commonly heard complaint is that supervisors are not willing to serve as role models nor reward the behavior that trainees have learned. It is fairly obvious that behavior learned in basic training, but not supported on the job, may not remain in the learner's repertoire. However, it is likely that this situation occurs in many organizations, including the Air Force. A basic research laboratory would provide a vehicle for both the identification of those constraints which affect transfer and an opportunity to study the effects. For example, it is quite possible to study the learning of inter-personal behavior on one task and then the transfer of that behavior to another situation in the laboratory. In the second situation, it is possible to examine situations where the supervisor is not prepared to support or reward the behavior. It is important to understand the transfer effects in those situations and also to explore variables in the learning situation to determine what support variables, if any, could maintain the appropriate behavior. Another organizational constraint which affects transfer behavior is the lack of opportunity to practice the learned behavior once the person is transferred to the job. A study by Laabs, Panell, and Pickering (1977) developed performance-oriented tests to diagnose deficiencies in job performance for United States Navy missile technicians. This study found that performance deficiencies in maintenance skills were related to the fact that the technicians were unable to practice their skills because of the extreme reliability of the missile test and readiness equipment. If similar situations exist in the Air Force, the organizational analysis would suggest the importance of studying procedures in the basic learning laboratory that would support behaviors that might have to be maintained without actual practice over long periods of time.

2. Research Issues Concerning Criteria for Training Validity

In general, examination of the types of criteria employed to provide evaluation data is not reassuring. As discussed in the paper, few studies have bothered to measure learning, and on-the-job behavior. It is probably not unreasonable to suspect that these investigations did not consider criterion relevance and reliability. The question about how criterion relevance is determined is not easily answered. There should be a scheme for the tracing-out process from the needs assessment to the development of criteria. The rules of this process are unknown. Indeed, investigators are not really very sure what the input should be to maximize criterion relevance. Recently there has been increased attention to criterion-referenced measures that are dependent on an absolute standard of quality (Swezey, 1978), rather than norm-referenced measures that are dependent on a relative standard. The characteristics of criterion-referenced measures - including methods of content validity to determine relevance as well as the design of methods to determine reliability indices - need attention.

There is even great difficulty in determining what constitutes ideal performance and how it should be measured. As discussed above, Williges, Roscoe, and Williges (1972) note that there is little agreement on what constitutes ideal pilot performance, and the reliability of most pilot-performance grading systems is disappointing. Thus, even studies of original learning are difficult to complete, because there is little agreement on what constitutes terminal performance. Even in the area of simulators, where there have been extensive efforts, the establishment of appropriate criteria for complex performance is a serious issue demanding the attention of researchers. In addition, continuing questions concern the development of new criteria measures that will provide further information about the instructional process. Thus, reaction measures that only ask the trainee for general feelings about learning have not been judged particularly useful. However, it is possible that trainees are able to self-assess their learning of particular components of KSAs. Research is needed to determine whether learners can assess their own capabilities concerning how much they have learned. There are also questions concerning the relationships between these self-assessment indices and other measures of learning in the instructional environment. A basic learning laboratory would provide an important vehicle to explore these issues.

Conclusions

This paper has presented the author's views concerning a hierarchy in the establishment of training, performance, and intra-organizational and inter-organizational validity. One way for the training analyst to consider this entire process is to ask "What is the final goal?" If the final goal is the establishment of intra-organizational validity, then it is necessary to establish all of the components of validity that precede it (i.e., both training and performance validity).

The examination of the needs assessment processes and evaluation processes as related to each of the validity goals raises a number of interesting instructional systems issues that should be examined in a learning laboratory.

REFERENCES

Blaiwes, A. S., & Regan, J. J. (1970). An integrated approach to the study of learning, retention, and transfer - a key issue in training-device research and development (NAVTRADEVCEN 1H-178). Orlando, FL: Naval Training Device Center.

Briggs, L. J., Campeau, P. L., Gagné, R. M., & May, M. A. (1967). Instructional media: A procedure for the design of multi-media instruction, a critical review of research, and suggestions for future research. Palo Alto, CA: American Institutes for Research.

Catalanello, R. F., & Kirkpatrick, D. L. (1968). Evaluating training programs - the state of the art. Training and Development Journal, 22, 2-9.

Cronbach, L. J. (1973). Course improvement through evaluation. Teachers College Record, 64, 672-683.

Gagné, R. M. (1970). The Conditions for Learning. New York: Holt, Rinehart & Winston.

Gilbert, T. F. (1960). On the relevance of laboratory investigation of learning to self-instructional programming. In A. A. Lumsdaine & R. Glaser (Eds.), Teaching machines and programmed instruction. Washington, DC: National Education Association.

Goldstein, I. L. (1974). Training: Program development and evaluation. Monterey, CA: Brooks/Cole.

Goldstein, I. L. (1978). The pursuit of validity in the evaluation of training programs. In I. L. Goldstein (Ed.), Training methodological considerations and empirical approaches. Special Issue Human Factors Journal, 20.

Goldstein, I. L., & Buxton, V. (1982). Training and human performance. In E. A. Fleishman, & M. D. Dunnette, (Eds.), Stress and performance effectiveness. New Jersey: Erlbaum.

Goldstein, I.L., & Rosenberg, B. (1977). An evaluation of a computer-assisted instruction program. Unpublished manuscript, College Park, MD: University of Maryland.

Goodman, P. S. (1969). Hiring and training the hard core unemployed: A problem of system definition. Human Organization. 28, 259-269.

Gordon, M. E., & Cohen, S. L. (1973). Training behavior as a predictor of trainability. Personnel Psychology, 26, 261-272.

Kraut, A. I. (1975). Prediction of managerial success by peer and training-staff ratings. Journal of Applied Psychology, 60, 14-19.

Laabs, G.J., Panell, R.C., & Pickering, E.J. (1977). A personnel readiness training program maintainence of missile test and readiness equipment (NPRDC-TR-77-19). San Diego, CA: Navy Personnel Research and Development Center.

Miller, R. W., & Zeller, F. H. (1967). Social psychological factors associated with response to retraining. (Final report), West Virginia University: Office of Research and Development, Appalachian Center, U. S. Department of Labor.

Mindak, W. A., & Anderson, R. E. (1971). Can we quantify an act of faith? Training and Development Journal, 25, 2-10.

Rosenthal, R. (1966). Experimenter effects in behavioral research. New York, NY: Appleton-Century-Croft.

Salinger, P. D. (1971). Disincentives to effective employee training and development. Washington, DC: U. S. Civil Service Commission, Bureau of Training.

Swezey, R. (1978). Criterion-referenced measurement in performance. In I. L. Goldstein (Ed.), Training: Methodological considerations and empirical approaches. Special Issue Human Factors Journal, 20.

Williges, B. H., Roscoe, S. N., & Williges, R. C. (1972). Synthetic flight training revisited (Technical Report ARL-72-21/AFOSR-72-10). Savoy, IL: Aviation Research Laboratory.

CHAPTER 8

RESEARCH ISSUES

Irwin L. Goldstein
University of Maryland

This paper briefly summarizes the research issues presented in the preceding chapters. The authors have presented their own individual views concerning contemporary research and theory, as related to the formation of a research agenda for a learning research laboratory. Interestingly, these individual chapters result in many common themes. The purpose of this chapter is to summarize and describe those themes.

Acquisition and Retention Issues

Glaser's chapter focuses on the principle that a major objective of Air Force training should be to produce expert apprentices. He defines expert apprentices as people who, upon leaving training, have strong capabilities to profit from the experiences they will receive in their military careers. Many of these issues refer to transfer-of-training questions. What does the learner have to acquire in training to result in more effective learning on the job? However, Glaser then takes the reader back a step by saying that first it is necessary to understand what most persons must acquire in order to be ready to profit from their experiences. Glaser then provides a number of parameters that should be studied in a learning laboratory. Some of these are variables such as categorization ability, knowledge of principles, mental models, and automatization. Glaser further asks questions such as: How is knowledge categorized and what principles must be gained in addition to understanding the procedures required? In addition, there are issues about what conditions of instruction favor the development of automaticity in basic skills.

Gagné addresses issues similar to those of Glaser and asks how it is possible to continue to maintain skills that are not constantly being used. He notes that these issues suggest a whole set of research questions about the adequacy of learning and their effects on later performance. Gagné suggests that it is necessary to begin to understand how learning proceeds from initial acquisition of a procedure to later stages of adequacy. He also notes the need to understand composition, which refers to how certain parts of a procedure are consolidated and also automatized so that its performance does not interfere with another task concurrently being performed. Gagné not only emphasizes research on variables affecting skill retention but also brings attention to the point that when training courses are of lengthy duration, skills learned early may be forgotten by the end of the program. An examination of this type of training program might provide a perfect vehicle for a learning laboratory to study skill retention issues. Gagné also emphasizes research on relearning, noting that one way of treating the skill retention problem is to emphasize relearning procedures.

In his chapter, Goldstein notes that there are a number of situations where the extremely high reliability of the technical equipment results in persons not having the opportunity to perform a task except in emergencies when the equipment fails. He calls for the study of procedures which would support behaviors that might have to be maintained without actual practice over a long period of time. Goldstein adds that there has been little attention given to studying variables in the job situation which support learning that has taken place in training. One variable that he suggests should be examined is the effects of maintenance of learned behaviors in situations where the supervisor is or is not willing to serve as role model or reward the learned behavior.

Royer also focuses on similar issues. His paper discusses techniques that either activate relevant prior knowledge or establish a bridge between prior knowledge and material to be learned. For example, Royer suggests that mnemonic techniques and the use of analogies for instructional purposes appear to hold a great deal of promise. He also notes that while some of these suggestions (e.g., mnemonic strategies) have been examined in educational settings, there is virtually no research on the utility of these techniques in military training situations. Royer also notes that some techniques, such as the use of analogies in instructional settings, are just beginning to be researched and would require intensive investigations in a learning laboratory.

Shuell points out that most learning for persons in the Air Force is procedural in nature, as in repairing a piece of equipment. However, he notes that declarative knowledge such as learning a radio call sign is also important. Yet, there is not much information about the relationship between these two forms of learning. Shuell asks whether emphasis on the declarative knowledge early in the learning process might be helpful in learning the procedural tasks required in Air Force jobs. In examining this issue, Shuell also notes the importance of examining a number of instructional variables, such as visual mnemonics and behavioral objectives.

Individual Difference Issues

Payne noted the need to determine the relationship between aptitude and training time. The issues that he raises are captured by asking what the payoff is in training in recruiting individuals with increased aptitude scores. In addition, he notes the importance of learning how these relationships are affected by the nature of the skill being learned and the aptitude chosen for examination.

Shuell also emphasizes this question, pointing out that such concerns must emphasize the individualization of instruction as well as questions about individual differences of people. Shuell asks a whole set of interesting questions including "How do experts who perform Air Force jobs differ from trainees?" He suggests that by understanding such differences, it will be possible to gain insight to instructional interventions that will be important to the learning process. Similarly Shuell would also ask what differentiates the performance of "good" and "poor" learners at various levels of competence.

Goldstein also discussed the importance of individual differences research. He notes that there is very little information available to indicate what type of individual difference measure best predicts performance at various stages of learning on different tasks. This type of information could be very useful in determining when individuals in a training program should be moved on to the next learning task.

Gagné also discusses the use of individual difference methodology to validate aptitude test scores against job-component tasks designed for research in the learning laboratory. Thus, by investigating learning performance on relevant tasks, the laboratory also affords the opportunity to examine the predictability of various aptitude measures.

Needs Assessment Issues

Goldstein notes that training analysts remain uncertain about the appropriate input for the design of instructional programs. He indicates that needs assessment of jobs has focused on a number of different factors, including the determination of the tasks to be performed; the knowledge, skills, and abilities necessary to perform the tasks; and the behaviors necessary to demonstrate competent performance. However, there is little information available comparing the various systems as inputs, nor is there much information about the utility of descriptive characteristics (e.g., frequency, criticality) to provide information about: (a) tasks, (b) knowledge, (c) skills or abilities, or (d) performance. For example, a consideration of the kinds of parameters that are directly relevant to the Air Force might suggest different types of training strategies for skills and abilities that are frequently performed, as compared to those skills and abilities which are critical to performance but are rarely performed. Both Goldstein and Gagné stress the same issue when they note that the CODAP database includes ratings of task difficulty that are obtained by asking supervisors "How long does it take to learn the task?" They indicate that the measure needs to be validated by comparing it with direct measures of "time to learn" on the same task. As they both indicate, this would make possible some interesting comparisons with the task difficulty database.

Gagné also points out that there are many different approaches to the selection of job tasks which are representative. Research examining the representativeness of job components, using both the CODAP system and measures such as time to learn, would provide important information on how to establish procedures to select representative job components. Royer in his ISD model also emphasizes the needs assessment process. His model begins with transfer task analysis. The idea here is to analyze the domain of activities a trainee will engage in after training is completed. Royer indicates the activities can then be used to serve as indices of the ability of training to provide transfer. This type of model, based on needs assessment procedures to select transfer objectives, could result in the design of learning tasks that are much more likely to produce transfer onto the job.

Criterion and Measurement Issues

All of the papers stress the importance of research on criterion issues. Gagné notes that low reliability has often been an obstacle to the use of job performance measures. He suggests the need for further research on criterion-reference measurement and the development of appropriate measures of reliability. Goldstein makes the same point and also suggests that the development of content validity procedures to validate criteria is badly needed. He notes that there is great difficulty in determining what constitutes ideal pilot performance and how such performance should be measured. Even studies of learning are difficult to complete because there is little agreement on what constitutes terminal performance. Royer further attests to the need for this research when he indicates that assessment of knowledge specifically relevant to a particular course should be tested as a predictor of end-of-training performance. He then indicates that the key to evaluating this hypothesis is to devise measurement procedures that reflect performance of a relevant activity actually included in training. What types of measure will be most useful remains a question. Also, whether different types of measures will best predict performance at different points in training also remains a research question. As Gagné notes, widespread evaluation of job performance on site in the Air Force would be prohibitively expensive. Validation of both learning and job performance measures would be an important task for a learning laboratory.

The topics above present only a few of the research themes mentioned in many of the papers. It should be noted that there were many other interesting and important topics discussed in the

AD-A169 014 LEARNING RESEARCH LABORATORY: PROPOSED RESEARCH ISSUES 2\2
(U) UNIVERSAL ENERGY SYSTEMS INC DAYTON OH
I L GOLDSTEIN ET AL. JUN 86 AFHRL-TP-85-54

UNCLASSIFIED F33615-82-C-0006 F/G 5/10 NL individual papers. For example, Glaser presents some important research suggestions about teaching and management variables that should be investigated. Thus, the individual papers should be examined in order to gain an appreciation for the large number of important topics that could be researched in a learning laboratory. Also, of course, the summary statements contained in this paper do not do justice to the breadth of ideas expressed in the papers. Again, the papers must be read individually to accomplish that.

CONCLUDING REMARKS

Lt Col David L. Payne
Air Force Human Resources Laboratory

Dr. Goldstein has presented an excellent summary in this chapter of the major themes of the report contributors. As he noted, the research proposals are numerous and have merit - both in their contributions to the general body of knowledge in the learning/training research areas, and in their potential utility for improving current military training programs. I would like to conclude by providing a brief statement as to my opinion regarding "Given all this - where do we go from here?" In doing so, I provide the perspective of a research psychologist who is familiar with the various military training systems and requirements.

First, the studies proposed that are natural follows-ons to research now being conducted as part of the AFHRL Learning Abilities Measurement Program (Project LAMP) should be given high priority. As the Project LAMP results help us to identify the components of learning ability, the next step is to determine how these components are modified by training and what instructional conditions are best suited to the task. As an example, what instructional techniques enhance the learning capabilities of those who are deficient in working-memory capacity? Or, as noted by Dr. Glaser, what conditions favor automaticity in the acquisition of basic skills? These are just two of the numerous research efforts which could capitalize on the results of on-going LAMP studies focusing on individual differences in information-processing skills.

Second, as suggested by Dr. Gagné, further research should be undertaken concerning the retention of learned skills. Results from earlier Project LAMP studies indicate that the entering Air Force recruits vary widely in their learning abilities, and their retention appears to be highly related to initial learning rate. Some questions that should be addressed in this area include the following: What are the effects of skill acquisition rates on the retention of different types of information? What techniques can be used on the job to enhance/maintain the retention of learned skills (Goldstein)? What are the effects of different instructional techniques on retention (e.g., mnemonics)? What techniques are best suited for the retention of different types of knowledge (i.e., procedural versus declarative)? This stream of research could have great utility for the military services in determining the appropriate lengths and instructional modalities of various courses in consideration of desired student retention and relearning requirements.

Finally, several authors noted the critical need for further research on criterion measures. The military services' personnel research laboratories are now participating in a joint effort to develop reliable measures of job performance. Their first efforts will be aimed at developing surrogates of actual "hands-on" performance measures. The surrogate measures can then be used to validate the services' selection and training procedures in an economically feasible fashion. The products of their efforts could also provide end-of-course criterion measures that could be tested as alternatives to current measures. The following two questions could be addressed: What are the effects of using different instructional techniques on student performance using the different criterion measures? Is the same type of learning involved when using the different measures? An example for the latter question is: Learning required to pass a job performance test may be more procedural in nature than learning required to pass a written end-of-block test (declarative).

In conclusion, I would like to express my sincere appreciation to the authors of these papers. Their truly outstanding efforts have provided the foundation for at least a decade of valuable research in the learning and training areas.

END

DTIC

8-86